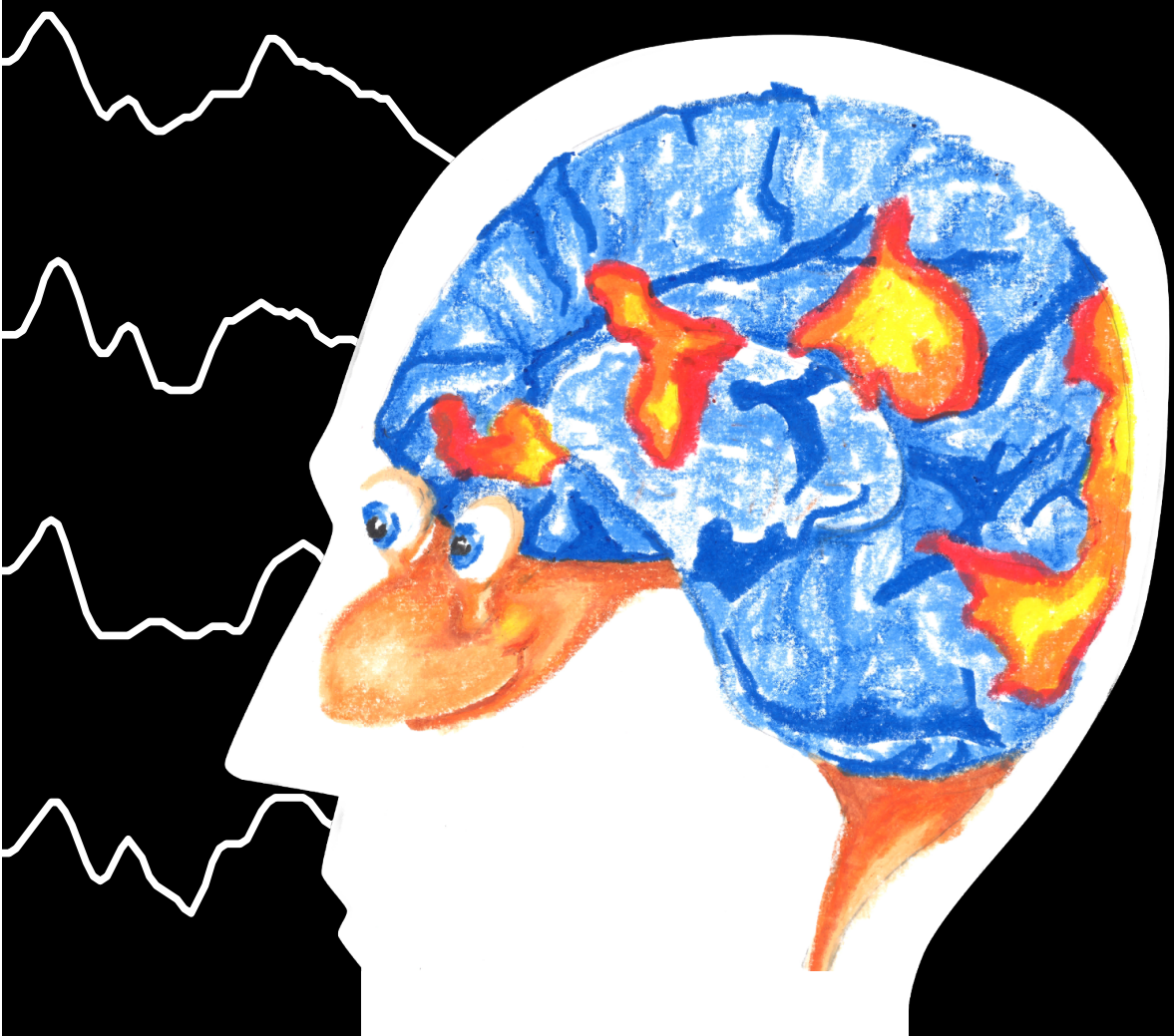
Studying semantic relationships using electroencephalography

doctoral dissertation of
Marijn van Vliet

Leuven, 2015



KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Neurosciences

KU LEUVEN

DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

STUDYING SEMANTIC RELATIONSHIPS USING ELECTROENCEPHALOGRAPHY

Wouter Marijn VAN VLIET

Jury:

Promoter: Marc Van Hulle
Chair: Jan Wouters
Jury members:
prof. Gerrit Storms
prof. Rik Vanderberghe
prof. Patrick Dupont
prof. Wim Fias
prof. Lars Kai Hansen

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

Oktober 2015

Acknowledgements

This thesis would not have existed if it weren't for the help of several people.

First and foremost, thanks to all my faithful subjects who willingly suffered reading through endless lists of flashing words. Twice.

I also owe a great deal to my promoter Marc Van Hulle. His willingness to embrace new research idea's led me to Leuven. It was he who found funding for it and continued to believe in me throughout the project. Early on in the project, I found my visits to Ghent to meet Roeljan Wiersema and Wim Fias very helpful. Their perspective from psychology and enthusiasm about the project were key in moving forward at the time. Special thanks to Simon De Deyne and Gert Storms, who's unrelenting data gathering resulted in word-class semantic norms for the Dutch language. These norms are a source of data that is as important to this thesis as my EEG recordings.

Furthermore, I would like to thank my coworkers Flavio Camarrone, Nikolay Chumerin, Adrien Combaz, Mansoureh Fahimi, Marie-Eve Joret, Elvira Khachatryan, Nikolay Manyakov, Anderson Mora Cortes, Arne Robben, Yelena Tonoyan and Benjamin Wittewrongel for their insights, proof reading manuscripts, being guinea pigs for pilot experiments, but most of all their friendship.

The cover of this thesis was designed by Dominique van Ipenburg. Thank you Dominique for always being by my side and Midas for being the best son a father could wish for; welcoming me back from the lab with laughter, hugs and kisses.

Marijn

Contents

Acronyms **V**

List of Figures **VIII**

List of Tables **IX**

1 Introduction **1**
1.1 Using semantic priming to study semantic memory . . . 1
1.1.1 Semantic priming studies . . . 1
1.1.2 Models of semantic memory . . . 2
1.2 Finding semantic relationships . . . 4
1.3 The N400 potential as a correlate of semantic priming . . . 5
1.3.1 The event related potential . . . 5
1.3.2 The N400 component . . . 8
1.3.3 Isolating the N400 through creating contrasts . . . 10
1.4 Multivariate filtering . . . 13
1.4.1 The linear model . . . 13
1.4.2 Data-driven filters . . . 14
1.5 Objectives . . . 16
1.5.1 Isolating the N400 component . . . 16
1.5.2 Dealing with confound variables . . . 17
1.5.3 Towards the construction of semantic clusters, based on N400 amplitude . . . 17

2 Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs **19**
2.1 Abstract . . . 19
2.2 Introduction . . . 19
2.3 Materials and Methods . . . 21
2.3.1 Subjects . . . 21
2.3.2 Ethics Statement . . . 21
2.3.3 Materials . . . 21
2.3.4 Experimental procedure . . . 23
2.3.5 Data preprocessing and extracting trials . . . 24
2.3.6 Statistical analysis . . . 24
2.4 Results . . . 24
2.4.1 Button responses . . . 25
2.4.2 ERPs during the speeded condition . . . 26
2.4.3 ERPs during the delayed condition . . . 27
2.4.4 Analyzing P3-latencies . . . 27
2.4.5 Impact of P3-latency on overall ERP . . . 31
2.5 Discussion . . . 34
2.6 Conclusion . . . 34

3 Single-trial ERP component analysis using a spatio-temporal LCMV beamformer 37
3.1 Abstract . . . 37
3.2 Introduction . . . 37
3.2.1 Limitations of the averaging technique . . . 38
3.2.2 Performance criteria for multivariate techniques . . . 39
3.2.3 The beamformer technique . . . 39
3.2.4 Assessment of various multivariate techniques . . . 40
3.3 Methods . . . 41
3.3.1 Linear model of EEG . . . 41
3.3.2 Multivariate filters . . . 43
3.3.3 Supervised learning approach . . . 44
3.3.4 Beamformer approach . . . 45
3.3.5 Modeling the activation pattern . . . 46
3.3.6 Evaluation on simulated data . . . 47
3.3.7 Evaluation on real EEG data . . . 49
3.3.8 Software . . . 53
3.4 Results . . . 53
3.4.1 Simulated data . . . 53
3.4.2 Real EEG data . . . 56
3.5 Discussion . . . 59
3.6 Conclusion . . . 60

4 Using the N400 to cluster words into categories 83
4.1 Abstract . . . 83
4.2 Introduction . . . 83
4.3 Methods . . . 85
4.3.1 Participants . . . 85
4.3.2 Materials . . . 85
4.3.3 Experimental procedure . . . 86
4.3.4 Data preprocessing . . . 87
4.3.5 Block modeling based on N400 amplitude . . . 87
4.4 Results . . . 88
4.5 Discussion and conclusion . . . 90

5 Discussion and conclusions 93
5.1 Summary of contributions . . . 93
5.2 Experimental paradigm and the P3 . . . 95
5.3 Simulation study comparing multivariate filtering approaches . . . 96
5.3.1 Unstructured noise . . . 96
5.3.2 Structured noise . . . 97
5.3.3 Temporal jitter . . . 99
5.3.4 Label accuracy . . . 99
5.4 EEG study comparing multivariate filtering approaches . . . 100
5.5 Practical considerations concerning the LCMV beamformer approach . . . 101
5.5.1 Designing a proper template . . . 101
5.5.2 Estimating the covariance matrix . . . 101
5.6 Clustering words . . . 102
5.7 The road ahead . . . 103

Summary 105

Samenvatting **107**

Curriculum vitae **109**

Bibliography **113**

Acronyms

b**CSP** bi-linear common spatial patterns.

f**MRI** functional magnetic resonance imaging.

l**SVM** linear support vector machine.

st**LCMV** spatio-temporal linearly constrained minimum variance.

AOA age of acquisition.

AG angular gyrus.

AIC Akaike information criterion.

ATC anterior temporal cortex.

BAS backward association strength.

BCI brain–computer interface.

CAR common average reference.

COI component of interest.

CSP common spatial patterns.

EEG electroencephalography.

EOG electrooculogram.

ERP event-related potential.

FAS forward association strength.

FC Fisher criterion.

GMM Guassian mixture model.

GOI generator of interest.

HAL hyperspace analogue to language.

HOOI higher-order orthogonal iteration.

ICA independent component analysis.

IFG inferior frontal gyrus.

IIR infinite impulse response.

JAM judgment of associative memory.

LCMV linearly constrained minimum variance.

LDA linear discriminant analysis.

LME linear mixed effects.

LR logistic regression.

LSA latent semantic analysis.

MEG magnetoencephalography.

MR motor related.

MRP motor related potential.

MTG middle temporal gyrus.

NS negative slope.

OAS oracle approximating shrinkage.

PCA principal component analysis.

REML restricted maximum likelihood.

ROI region of interest.

RPM random permutation model.

RT response time.

RVM random vector model.

SOA stimulus onset asynchrony.

STG superior temporal gyrus.

SVM support vector machine.

List of Figures

1.1 The spreading activation model . . . 3
1.2 Example ERP recording in response to visually presented words . 7
1.3 MEG activation in response to congruent versus incongruent sentences . . . 11
1.4 Schematic of the functional neuroanatomic model for semantic processing proposed by Lau et al. . . . 12
1.5 The effect of the amplitude of an ERP component in the linear model 14

2.1 Forward association strength of all 800 word-pairs that were used 22
2.2 Boxplots of button responses made by the subjects . . . 25
2.3 Stimulus-locked ERPs and scalp topographies for both experimental conditions . . . 27
2.4 Response-locked ERPs and scalp topographies for both experimental conditions . . . 29
2.5 Two-dimensional histogram of single trial component latency versus RT . . . 29
2.6 Grand average ERPs demonstrating the effect of differences in P3-latency . . . 32

3.1 Sketch showing how signals generated at different dipole generators are captured by the EEG electrodes . . . 42
3.2 Pearson's correlation coefficients between dependent variables used in this study . . . 50
3.3 Performance of various techniques to estimate component of interest (COI) amplitude under various simulated conditions . . . 53
3.4 Grand average ERP in response to word-pairs with increasing FAS during the delayed condition . . . 56
3.5 Univariate regression between each analysis method and each stimulus property . . . 58
3.6 Temporal (left) and spatial (right) activation patterns used as template for the COI during beamformer analysis . . . 59

4.1 Estimated N400 amplitude for each word-pair . . . 88
4.2 Optimal clustering by the block model . . . 89
4.3 Scores for all possible ways to cluster the words into two groups . 90
4.4 Grand average ERPs . . . 91

List of Tables

2.1 Descriptive statistics of the bins across all conditions 33

3.1 Mathematical notation 41
3.2 Stimulus properties used in the regression study 51

4.1 Words used in the clustering experiment 86

Chapter 1

Introduction

1.1 Using semantic priming to study semantic memory

The semantic system in our brain gives us the capacity to exchange messages with others at high speed, which is extremely useful in our day to day activities. Semantic priming experiments [1, 2] have revealed one of the manners in which the semantic system optimizes language processing. These experiments found that, since semantic tokens[1] usually occur in a logical sequence (i.e. a story), the semantic system extrapolates from previously processed tokens to facilitate the processing of continuations that are more likely [3]. This process is referred to as the priming effect, where processing of a *target* token can be the facilitated by the preceding context, which is referred to as the *prime*. The priming effect demonstrated that our semantic system is closely coupled to memory systems. Further research has linked it to our implicit (procedural, non-declarative) semantic memory [4], which provides unconscious recollections of world-knowledge[2]. This makes priming studies a popular tool not only to study aspects of the semantic system, but also, through careful experimental design, aspects of our semantic memory [3, 6, 7].

1.1.1 Semantic priming studies

The semantic priming effect has been shown to be robust across a variety of stimulus modalities, for example written or spoken words, organized in sentences or as word-pairs [8–10], environmental sounds [11], line drawings [12] and animations of action sequences [13]. Priming has classically been demonstrated by measuring the response time (RT) of a subject during a task that requires making a decision based on the meaning of a stimulus [3, 14]. The stimulus for which a decision is required is referred to as the *target* and the context preceding the target is refereed to as the *prime*. A commonly used task is lexical decision [8]: the subject is presented with a letter string that is either a valid word (NURSE) or not (NURFE). Faster and more accurate responses are obtained for valid words when they were preceded by a semantically associated word (DOCTOR) [15]. Other tasks can be employed as well, for example:

- Reading aloud, where the subject pronounces each word [16]

[1] In this thesis, the term *semantic token* is used to refer to the underlying meaning of a word. For example, the semantic token can be activated by the written or spoken word DOG, as well as a picture of a dog or an encounter with a live dog.

[2] In contrast, explicit (declarative) memory is employed when we consciously recall a specific event, fact or rule [5].

- judgment of associative memory (JAM), where the subject is presented with two stimuli and asked to judge whether on average, people would say they were related or not [17]

- Letter decision, where the subject is asked to fill in a missing letter [18]

As the subject strives to optimize performance, the semantic system tunes itself to the task at hand. Therefore, the rate at which various properties of the stimuli affect semantic priming is dependent on the chosen task [19], as well as the manner in which the stimuli are presented, such as the stimulus onset asynchrony (SOA) and the probability that the prime and target are congruent [20].

1.1.2 Models of semantic memory

It is thought that the relationships between semantic tokens are a key element in the organisation of our semantic memory. A popular model that forms the basis for many semantic priming experiments is the spreading activation model [21, 22] (figure 1.1). In this model, whenever a semantic token is activated in semantic memory, related tokens become pre-activated. These pre-activated tokens activate other tokens in turn, causing activation to spread throughout the network-like structure. Memory access to these pre-activated tokens is facilitated, although this facilitation decreases with every step away from the token that was originally activated.

This model can account for many priming effects, notably the mediated priming phenomenon [23–25]. In mediated priming, the prime and target do not share a direct associative connection, but there is an indirect path through an intermediate token (LION–STRIPES, BULL–MILK). The indirect path produces a priming effect, although it is weaker than in the case of a direct link [26]. Studies with masked priming, where the prime is presented for such a short duration that the subject is not consciously aware of it, demonstrated the automatic nature of the priming effect [27–31].

While the spreading activation model is a useful tool for interpreting the results of semantic priming studies, it is not necessarily an accurate model of how semantic tokens are represented in our semantic memory at a lower level. For example, some recent models do not assume that semantic tokens have a local representation in the brain, with links between them, but rather model them as a distributed pattern of activation across multiple brain regions [32–34]. In such models, tokens with similar meanings have similar distributions and spreading activation is found to be an emergent behaviour of such a distributed system [35].

If one uses a network model to describe semantic memory, the question arises what the nature is of the relationship between the tokens. It has been observed that many different kinds of relationships can contribute to the priming effect. For example, studies have found effects on priming for associative relationships [8–10] as well as feature overlap [19, 36, 37], semantic category membership [38] and rhyming [39]. Although it is likely that multiple kinds of relationships play a role [38], there is some debate about the exact contribution of

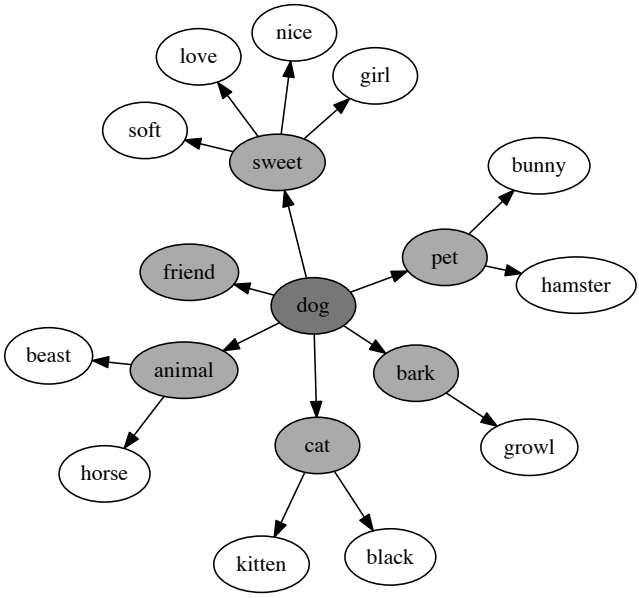


Figure 1.1: The spreading activation model. When a semantic token (here: DOG) is activated in semantic memory, related tokens are pre-activated as well. These pre-activated tokens activate other tokens in turn. With each step, the activation level grows weaker, indicated here by fading background shading.

each relationship type [25]. Taken as a whole, our semantic system seems apt at exploiting almost any kind of relationship between semantic tokens that carry some predictive quality, given the task at hand.

1.2 Finding semantic relationships

Regardless of whether semantic relationships play a role at the low or high level organization of semantic memory, they are regarded as a vital component in how we acquire, forget, recognize and recall semantic tokens [1, 3, 40]. Therefore, a large body of work is dedicated to finding them.

One way to obtain semantic relationships for a large collection of word pairs is to perform word co-occurrence analysis on a large text corpus. Generally, the type of relationship (e.g. syntactic or semantic) extracted is dependent on the algorithm used. For example, the hyperspace analogue to language (HAL) model [41] counts the number of co-occurrences between words with a fixed window size. More complex measures include latent semantic analysis (LSA), which treats words that co-occur in the same document as related [42]. Random vector models (RVMs), such as the random permutation model (RPM) [43], assign an initially random vector to a word. Each time the word is encountered in the corpus, its vector is updated based on the vectors of the surrounding words. In this manner, the vector representations of words that tend to co-occur become more similar. Another approach would be to use dictionaries and thesauri, which is one of the foundations of the WordNet database [44].

To obtain a measure that is more closely tied to associative relationships, one can employ a free association task. The semantic relationship between two words, the *prime* and *target*, has a notion of strength, which can be expressed by a variety of measures. For example, even though a semantic relationship between KITCHEN and SPOON exists (spoons are regularly found and used in kitchens) it might be weaker than the relationship between KEY and LOCK (keys are almost always used in combination with a lock). The forward association strength (FAS) measure is obtained by asking a large population of participants to write down the first word (or first few words) that come to mind when presented with a cue word [45–47]. The FAS is the proportion of respondents that wrote down the second word in response to the first [48]. Likewise, backward association strength (BAS) is the proportion of the respondents that wrote down the first word in response to the second.

The advantage of the methods listed above is that they allow a researcher to extract and use semantic relationships on a large scale. However, the question remains how well these relationships correspond to the actual relationships that give rise to the semantic priming effect and hence play a role in the organization of our semantic memory. Therefore, it will be useful to employ the semantic priming paradigm itself to map out semantic relationships.

Given that semantic relationships are employed in the brain even during early stages of semantic processing, before conscious awareness [28–31], it is useful to obtain a measure of relationships that is based on an automatic response. Also, the aforementioned methods are based on data gathered from thousands

of individuals, thus representing the great common denominator. It would also be interesting to be able to study the networks of not so common individuals, such as those diagnosed with autism, aphasia, Alzheimer's or schizophrenia, which brain responses to semantic stimuli are known to be anomalous [49].

In this thesis, an attempt is made to obtain a measure of semantic relationship strength through analyzing recordings of brain activity in a semantic priming setting.

1.3 The N400 potential as a correlate of semantic priming

In order to capture brain activity, this thesis employs electroencephalography (EEG), which is a non-invasive method that is widely used in the neurosciences to study ongoing processes in the living brain. In EEG, electrodes are placed on the scalp that measure the minute changes in electrical potential produced by post-synaptic neural currents [50].

The electrical potential is recorded as the difference between two electrodes, in volts. Generally, a measurement is performed between each electrode and a reference electrode, which is determined by the EEG hardware. The resulting signal can than be re-referenced through numerical computation to use any combination of electrodes as new reference. It is advisable to place one or more reference electrodes in locations that are considered to provide a good contrast with the location where the main effect is expected. Common referencing areas are the nose, earlobes or mastoids [1]. Optionally, one can apply a common average reference (CAR) by subtracting the channel-wise average from each individual channel [51]. The choice of reference is important as it can greatly influence the shape of the EEG waveforms (however, not its spatial distribution [51]) and hence effect size. All the recordings performed in this thesis were done on hardware that had its reference electrode in a central-parietal position (near the Pz electrode). The signal was later re-referenced to the mean of two mastoid electrodes.

While performing an EEG is a non-invasive and relatively unobtrusive procedure and offers excellent resolution in time[3], one of its downsides is that it depends on the flow of electric current travelling through the headwards the electrodes. Along the way, the signal spreads out throughout the head due to volume conduction, with causes each electrode to pick up signals from practically all parts of the brain [52]. This means that EEG has a poor spatial resolution and always records a mixture of various brain processes that happen to coincide in time [53]. If one is interested in studying one particular brain process, as is usually the case, these signals must be separated in order to isolate the process of interest.

1.3.1 The event related potential

A commonly used technique to do so is the construction and analysis of the event-related potential (ERP). During ERP analysis, the subject performs a cer-

[3] the maximum sampling rate of the hardware used in this thesis is 2048 Hz

tain task multiple times, referred to as trials, and the brain responses for each trial are averaged [54]. Signal components that differ between trials tend to cancel out and signal components that are similar between trials (the event-related part of the signal) remain. This leaves a waveform which peaks and valleys are attributed to various brain processes that are evoked by the stimulus, referred to in this thesis as ERP *components*.

The naming of these components generally follows the convention of indicating whether it is a positive (P) or a negative (N) deflection, along with the typical latency in milliseconds, relative to the onset of some event, such as the presentation of a stimulus. However, since the direction of the deflection (positive or negative) varies according to the choice of reference electrodes and the timing of components can vary according to the experimental paradigm, different studies sometimes assign different names to what is essentially the same component. Alternative naming schemes, such as counting the number of peaks or valleys leading up to the components, suffer from similar disadvantages.

ERP components also have an *amplitude*, a term which is used to indicate the overall size of the peak or valley in the ERP waveform. For example, in figure 1.2, the N400 component has a larger amplitude during the condition where the prime and target words were unrelated, compared to the "related" condition. This amplitude is sometimes quantified as the peak amplitude of waveform in a certain time interval. More commonly, the mean amplitude during a time interval is used [55]. Sometimes, the amplitude is measured relative to the mean ERP voltage during a pre-stimulus (e.g. −100 ms to 0 ms) time interval. However, it is also common to measure the effect of an experimental manipulation, in which case the difference in component amplitude (either as peak amplitude or mean across a time window) is measured between experimental conditions. For more information about recording and analyzing ERPs, see Luck (2005) [54].

In the case of visual word processing (figure 1.2), the following landmark components are generally observed:

N100 Also referred to as the N1, or N/P150. This is a component that shows a dipole pattern across the scalp (figure 1.2, first scalp plot) and is one of the earliest brain potentials sensitive to changes in attention [56]. In the context of visual word processing, it is associated with early visual processing concerning the detection of letter shapes [57, 58]. At this stage, the signal is sensitive to the location of the word in the visual field of the subject and the shape (i.e. the font) of the letters [59].

P250 Also referred to as the P2, P200, P250 or the "recognition potential" [60]. This component has been linked to selective attention [61], as well as access to short-term memory [62, 63]. In the context of visual word processing, it is also associated with the mapping of orthographic information to a whole-word representation [30, 58]. This component has been shown to be sensitive to repetition priming [60], where the size of the component correlates with the amount of orthographic overlap between the prime and target stimulus [30]. In figure 1.2, it is likely that the P250 overlaps with an P3a component, which will be discussed next.

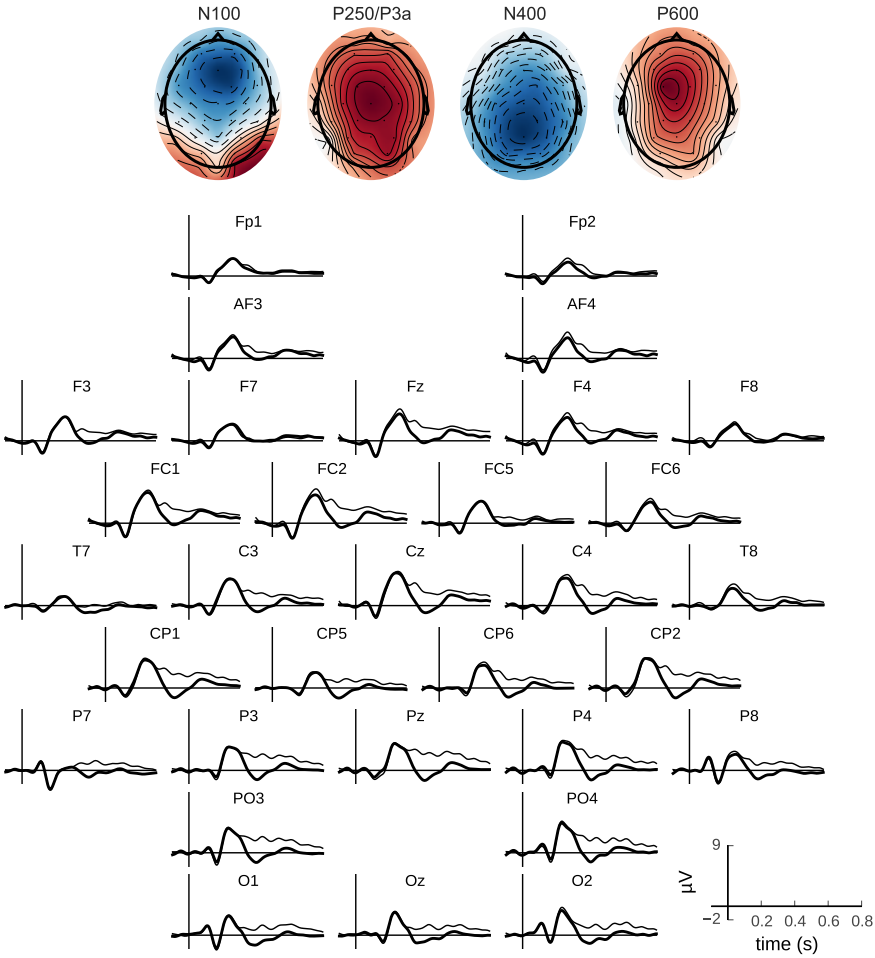


Figure 1.2: Example ERP recording in response to visually presented words. The ERPs plotted in this figure were constructed by averaging across 8400 trials, collected from 33 subjects. The timing is relative to the onset of presentation of the target word. The stimuli were word-pairs with either high FAS (thin line) or low FAS (thick line), which were shown using sequential visual presentation. The task given to the subjects was to determine whether the prime and target words were related or not. Subjects delayed their button response with 1 s. For more details about the methods used to produce this figure, we refer to section 2.3. Scalp topographies are shown (top) for several landmark components. For the N400 topography, the difference between the two conditions is shown at 400 ms; the other topographies show the voltage relative to zero at the indicated time.

P300 Also referred to as the P3, this component has been extensively studied, most notably using the oddball paradigm. In the oddball paradigm, the subject is asked to respond to an "oddball" stimulus, which is a rare event embedded in a sequence of common events. Interestingly, the denominators "rare" and "common" can be assigned based on arbitrary features of the stimulus. For example, P300 brain–computer interfaces (BCIs) operate by presenting all the letters of the alphabet with equal probabilities. The user can make one of the stimuli a rare event by mentally selecting the corresponding letter as a target, making the other letters frequent non-targets. This causes the mentally selected letter elicit a larger P300 component when presented than the non-target letters [64–66]. While a binary decision between two events with equal probability is sufficient to elicit the P300, its amplitude correlates negatively with the probability of the target stimulus [67].

A distinction is made between early P3 (P3a) and late P3 (P3b) [68]. The P3b is elicited when a subject is asked to discriminate between stimuli. In contrast, the P3a is elicited by attention grabbing stimuli that do not require action by the subject or by "no-go" distractor stimuli [69]. In figure 1.2 the P3a component is likely overlapping with an P250 component. In chapter 2, the role of the P300 in semantic priming experiments is discussed more in-depth.

N400 The N400 (figure 1.2, third scalp plot) is a component that occurs when semantic memory is accessed in order to translate the word-shape into a semantic token that carries meaning [7]. Its amplitude is closely tied to the ease of this memory access and therefore closely tied to the semantic priming effect [70]. As the main component of interest in this thesis, it will be discussed in more detail in section 1.3.2.

P600 The P600 (figure 1.2, last scalp plot) is shown to be mostly sensitive to syntactic errors, such as using the wrong pronoun or tense of a verb in a sentence [71, 72]. However, it has also been shown sensitive to semantic violations, if the detection of the violation requires interpreting the message conveyed by multiple stimuli [73]. Where the N400 is predictive of the ease with which the current stimulus is processed, the P600 is rather predictive of the ease of integrating multiple stimuli in a coherent message [28].

For a more in-depth discussion on the ERP components occurring during semantic processing, see Kutas et al. [74]. The focus of this thesis lies mainly on the N400 component.

1.3.2 The N400 component

The N400 component (figure 1.2, third scalp plot) was first described by Kutas and Hillyard [75]. In an experiment using sentences that either ended congruently or not, they sought to produce a linguistic oddball effect that may elicit a P3b component. However, they found a new potential which they named the

N400, as it was a negative going wave that peaked around 400 ms. They found that the amplitude of the N400 in response to the sentence final word correlates strongly with its cloze probability: the percentage of people that identify (in a questionnaire) the word as the most probable conclusion of the sentence. The N400 amplitude was larger for sentence final words with a low cloze probability (I drink coffee with milk and *socks*) relative to those with a high cloze probability (I drink coffee with milk and *sugar*).

In a follow up experiment, Kutas and Hillyard demonstrated that the N400 could also be modulated by presenting prime–target word-pairs [76]. The analogue to cloze probability for word pairs is the FAS measure, which was shown to correlate with N400 amplitude in much the same way cloze probability did in sentences. Target words preceded by a prime with a low FAS (submarine–carpet) produce a larger N400 amplitude than those preceded by a high FAS prime (locomotive–train). This result is reproduced in chapter 2 (figure 2.3).

In the wide range of semantic priming experiments that followed, the N400 has been shown to correlate with many factors that influence word processing speed [1, 7, 77]. Some of these factors are properties of the target stimulus, for example:

- Frequency of use [78]
- Number of orthographic neighbours [79]
- Age of acquisition (AOA) [80]
- RT to the word in isolation [81]
- Concreteness rating of the word (CHAIR is a more concrete term than LOVE) [82]

But more importantly, other factors are properties of the semantic relationship between the target word and a preceding prime word, for example:

- FAS [83]
- BAS [77]
- World knowledge [84]

All of the above have also been observed to influence RT in semantic priming experiments [85].

However, an N400 effect is not obtained for all unexpected manipulations using words [7]. For example, the original experiment of Kutas and Hillyard also included a deviant stimulus size condition (I shaved off my mustache and BEARD), which was found to elicit a later positive component and not an N400 [75]. Furthermore, not all linguistic manipulations that affect RT in semantic priming experiments necessarily affect N400 amplitude. For example, errors in grammar (All turtles have four leg) elicit an P600 rather than N400 [71, 72] and false statements that contain a negation (A robin is not a bird) do not necessarily affect the N400 (but can) [86]. Increasing the word length (the number of letters)

increases RT, but mostly affect the amplitude and latency of earlier (N100, P250) components and not the N400 [87], at least, when properly disentangled from the correlated word frequency metric, which does affect the N400 [78].

Similarly to RT, the semantic priming effect has been shown to modulate N400 amplitude across different modalities, such as line drawings [11, 88] and environmental sounds [11, 89]. This indicates that at least some aspects of the semantic memory are amodal. However, although there are broad similarities in wave shape and timing in the obtained N400 effects, there were also important differences in scalp topography. For example, pictures elicit a more frontally distributed N400 [88] than written words, and familiar faces elicit a more occipital N400 effect [90]. These studies point to a semantic system that responds of which some aspects are amodal, but also contains distinct neural areas that specialize in different modalities [7].

The neural origins of the N400 have been traced to a network of area's [91, 92]. In magnetoencephalography (MEG) studies, activations in semantic priming studies that resemble the N400 are mostly seen in the middle temporal gyrus (MTG) and superior temporal gyrus (STG), with additional activations in the inferior frontal gyrus (IFG) [93–95] (figure 1.3). Reviewing both MEG and functional magnetic resonance imaging (fMRI) results, Lau et al. [92] proposed a network structure where different cortical area's are involved in the activation of semantic tokens, their integration with the ongoing context, and top-down mediation between candidate tokens (figure 1.4).

From a functional standpoint, it is interesting that the N400 is both modulated by stimulus properties that presumably play a role during the early stages of visual word processing, such as orthographic neighborhood and word frequency, and properties that play a role at a later stage, such as dependence of the context of the entire sentence. As such, it flies in the face of models that treat word processing in a neat, sequential manner, where semantic context effects can only play a role after successful recognition of the word. The N400 seems functionally situated at the nexus where low-level word features meet the high-level semantic context [6, 7, 97].

Collectively, the results of N400 studies show a pattern where the amplitude of the N400 decreases with the ease in which visual and auditory features can be translated into a meaningful semantic token. This is affected by many properties of the target stimulus but, importantly, also the semantic relationship between it and the prime. Therefore, the N400 amplitude could serve, under the right circumstances and with careful counterbalancing of the stimuli, as a proxy measurement of the strength of the semantic relationship between words as utilized by the semantic system.

1.3.3 Isolating the N400 through creating contrasts

In order to use the N400 as a measure of relationship strength between words, this component will need to be isolated from the rest of the EEG signal.

The classical way to isolate a single ERP component, the COI, from other nearby and possibly overlapping signals, is to create a contrast. The goal of such

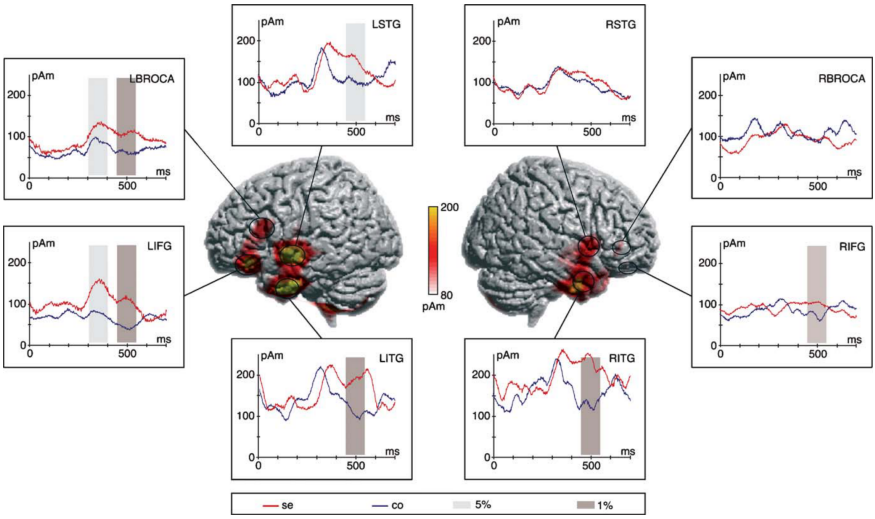


Figure 1.3: MEG activation in response to congruent versus incongruent sentences. Views onto the left and right hemispheres displaying the mean current distribution of the semantically incorrect condition at a latency of 350 ms. The mean time courses for each of the specified regions of interest (ROIs) and the two experimental conditions are displayed in boxes: semantically incorrect (red) and correct (blue). The ROIs are visualized by black outlines. Gray bars indicate that the t-test for this time interval yielded a significant difference between conditions ($p < 0.05$: light gray; $p < 0.01$: dark gray). Figure reproduced from [96] with permission.

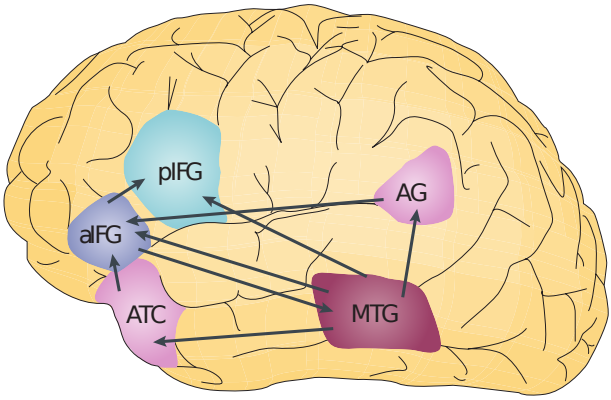


Figure 1.4: Schematic of the functional neuroanatomic model for semantic processing proposed by Lau et al. Representations are stored and activated in the middle temporal gyrus (MTG) and in the nearby superior temporal sulcus and inferior temporal cortex, and are accessed by other parts of the semantic network. The anterior temporal cortex (ATC) and angular gyrus (AG) are involved in integrating incoming information into current contextual and syntactic representations. The anterior inferior frontal gyrus (IFG) mediates controlled retrieval of lexical representations based on top-down information, and the posterior IFG mediates selection between highly activated candidate representations. Figure reproduced from [92] with permission.

a contrast is to construct two or more experimental conditions that generate the COI with a different amplitude across conditions. At the same time, the experimental conditions should be constructed in such a way that the amplitudes of any other ERP components that are generated are the same across conditions. Generally, it is only possible to create conditions in which the amplitudes of ERP components are similar *on average*, so averaging across many trials is a necessity. With a proper contrast, one can focus on the differences in the ERPs, generated in response to each condition, by creating the difference-wave between them. If the contrast has been properly constructed, the difference-wave should capture mostly the COI.

The scalp distribution and timing of the N400 was originally defined by the difference wave in a contrast between high and low sentence cloze probability [70]. Therefore, this contrast can by definition be used to isolate the N400 component.

In this thesis, two contrasts are used to isolate the N400. In chapters 2 and 3, the contrast used to isolate the N400 is an experimental condition where word-pairs are shown with a low FAS, versus a condition where pairs are shown with a high FAS. This contrast has been shown before to capture the N400 [12, 36, 83, 98–101]. For example, in figure 1.2, which uses the low FAS versus high FAS contrast, the component indicated as the N400 peaks around 400 ms and shows a posterior scalp distribution that is typical for the N400 [7]. In chapter 4 a contrast is used where in one condition, the prime and target word belong to the

same semantic category, versus the condition in which they belong to different ones. Is is another contrast that has been shown to capture the N400 [6, 99, 102, 103].

1.4 Multivariate filtering

If we wish to use ERPs to obtain a measure of strength between two words, we will have to obtain N400 measurements for this specific stimulus. However, repeating stimuli to create an ERP opens up an array of confounding repetition priming effects on the N400 [98], although the effect of semantic relationships remains [104, 105]. This means a method is desired to obtain accurate N400 amplitude measurements with as little averaging across stimuli as possible.

The need for averaging in ERP studies stems from the fact that the signal of interest can be generated deep in the brain. As it spreads outwards, it mixes first with nearby signals and later even with remote signals if the distance to the electrodes is large enough. Averaging across multiple trials is an effective way to reduce the influence of task-irrelevant signals and other noise. However, when one wishes to avoid averaging as much as possible, other ways of isolating the signal of interest must be explored.

1.4.1 The linear model

Since the mixing of signals inside a volume conducting medium such as the human head, occurs in a linear way, linear de-mixing strategies are among the most successful alternatives to averaging [106]. One of the advantages of linear methods over more complicated and powerful non-linear ones, is the lower number of parameters that need to be estimated, which translates into less data being required to produce a stable model. In this thesis, an algorithm that applies a linear transformation to the data in order to isolate a certain signal will be referred to as a linear filter.

Assuming that the recorded signal is the addition of the ERP COI with interfering signals:

$$\mathbf{X} = \mathbf{S} + \mathbf{N} \tag{1.1}$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$ contains the EEG recording during a single trial, recorded at m electrodes at n time samples, $\mathbf{S} \in \mathbb{R}^{m \times n}$ contains the EEG activity attributed to the COI, and $\mathbf{N} \in \mathbb{R}^{m \times n}$ contains everything else that is not the COI. Note that $\mathbf{N}$ not only includes unstructured, zero-mean noise, but also unwanted signals with clear spatial and temporal distributions, such as eye movement artefacts and signals produced by brain processes that are irrelevant for the study. The linear model assumes that when the amplitude of the COI increases, this causes a linear scaling of $\mathbf{S}$:

$$\mathbf{S} = y \cdot \mathbf{A} \tag{1.2}$$

This amplitude is denoted y in the model and can vary from trial to trial. Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is referred to as the activation pattern. See figure 1.5 for a sketch on how different values of y manifest in the EEG signal.

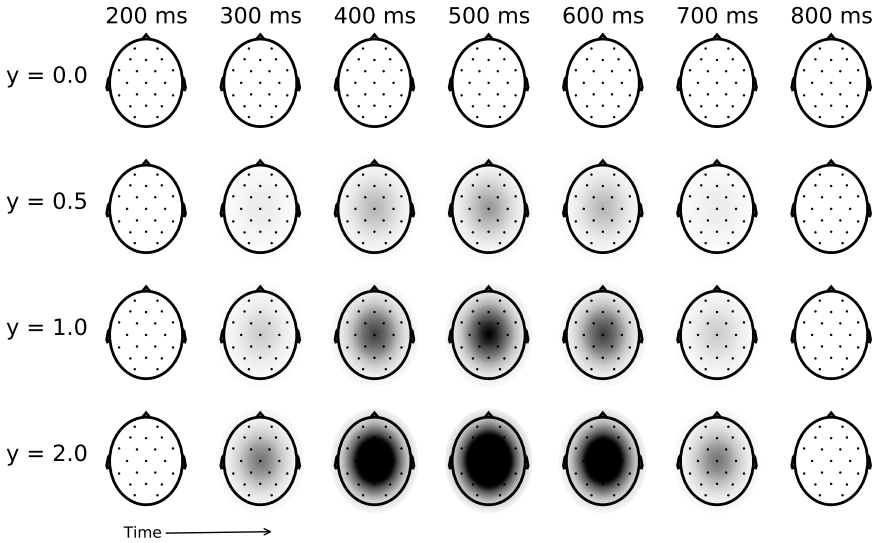


Figure 1.5: The effect of the amplitude (y) of an ERP component in the linear model. One row represents the EEG signal recorded during a single trial, in the absence of any other signals but the ERP component (**S**). The component has an activation pattern that is distributed both spatially and over time. This activation pattern is linearly scaled by the amplitude.

A linear spatial filter combines the signal from all channels into one (or several, but generally $< m$) virtual component. Different spatial filters generate components with different characteristics. For example, a spatial filter might aim to produce a component which optimally reveals experimental effects [107], or it might aim to extract the activity originating from a specific patch of cortex [108]. A linear spatio-temporal filter takes it one step further and combines the signal from all channels and all time points into one value that represents the entire trial.

For the purposes of this thesis, we wish that value to represent y, namely the amplitude of the COI, specifically, the amplitude of the N400.

1.4.2 Data-driven filters

In order to construct a linear filter that deduces y from a given **X**, estimations need to be made of both **A** and **N**. As soon as either changes, the filter needs to be recalibrated. Since the EEG will differ between recording sessions (e.g. some electrodes are in a slightly different place, make better or worse contact with the scalp, the recording is done on a different subject, etc.) and even within a recording session (e.g. electrodes might shift, the subject starts sweating, the subjects's mental state changes, etc.) a filter needs to be able to adapt. This means many successful filters have a data-driven component.

One can design a purely data-driven filter. Blind source separation meth-

ods do not have any notion about the signal of interest and are completely defined by the data. For example, principal component analysis (PCA) decomposes the recording into components that are uncorrelated with each other, while independent component analysis (ICA) decomposes the signal into components that are statistically independent [109]. These methods can be used as a spatial filter by selecting one (or a few) components and dropping the others. An example of a purely data-driven spatio-temporal filter would be a clustering approach like a Guassian mixture model (GMM) [110]. The effectiveness of these methods for the purposes of this thesis depends greatly on whether the N400 is in fact uncorrelated or statistically independent from the interfering signals and whether it happens to be neatly captured in a single component or cluster. When this is not the case and the N400 activity is spread out among multiple components, the best option will be to select components with the highest ratio of N400 versus interfering signals.

A better result might be obtained by using a filter that takes some information about the signal of interest into account. Supervised learning algorithms use training labels in addition to the EEG data to estimate both **A** and **N**. Linear regression models in particular have often been shown to be effective spatio-temporal filters for EEG signals [111–113]. These algorithms use multiple sets of recordings. First, a set of recordings that are labelled with the desired output is analysed: the training examples. The algorithm tunes the filter to reproduce the desired output from the training examples. Once estimated, the filter can then be reused to process other recordings. One way to structure the training examples is to have both recordings that contain a strong N400 activation and recordings that do not [114]. For example, by presenting word-pairs with a strong and weak FAS. From such a training set, the filter will be tuned to optimally distinguish between the two cases. If the only discriminating property between the cases is the presence of the N400, the filter will effectively isolate it. Examples of spatial filters that employ supervised learning are common spatial patterns (CSP) [115] and xDAWN [116], among others [117]. Examples of spatio-temporal filters that employ supervised learning are classifiers such as linear discriminant analysis (LDA) and linear support vector machines (lSVMs) [113], as well as spatio-temporal extensions to CSP such as bi-linear common spatial patterns (bCSP) [118].

Supervised learning algorithms can be very effective filters for isolating ERP components [119], however, it is generally difficult to obtain reliable training labels. The most common way to produce training labels is to run a training session with the subject. This means that if we desire a filter that isolates the N400 component, we face the difficult task of evoking an N400 response with an amplitude that is known a-priori.

Furthermore, supervised learning algorithms estimate their internal representations of both **A** and **N** from the data. Estimations for these can be difficult due to the high number of parameters involved. For example, a spatio-temporal filter designed to filter a one second segment of EEG recorded at 32 electrodes at a conservative sample rate of 50 Hz, contains 1600 parameters. In order to produce a stable filter, either a large amount of training examples need to be

collected, or strong regularization must be applied [112, 113].

1.5 Objectives

The major challenge addressed in this thesis is to obtain an accurate metric of relationship strength between two words, by measuring the amplitude of the N400 of a subject in response to the words in a semantic priming paradigm. Specifically, two roadblocks are tackled that hinder the mapping between a measurement on the ERP waveform and relationship strength:

1. Isolating the N400 component from other signals, such as sensor noise, other ERP components and other EEG
2. Dealing with the many confound variables that influence N400 amplitude, other than prime–target relationship strength

1.5.1 Isolating the N400 component

Chapter 2 explores the issue of overlapping ERP components. In classic semantic priming paradigms, the subject is asked to respond to the stimuli as fast and accurately as possible. The obtained RT measurements are then analyzed to quantify the priming effect. When the field moved from RT measurements to include EEG as well, the subject task should have changed, as now the moment of pressing a button generally falls within the time window used for ERP analysis. In this chapter we demonstrate the complications by asking subject to perform a JAM rating task (are the prime and target related?) using both a speeded response scenario and one where delayed their button response. We show how a large P3b component interferes with the N400, which, if not dealt with properly, will lead a large confounding effect of RT on the estimated N400 amplitude. While this effect has been noted in early studies in passing [120, 121], many modern N400 studies still employ a speeded button response task without acknowledging the confounding P3b component[4]. Delaying the button response of the subject alleviates the influence of the P3b.

Although some interfering signals, such as overlapping ERP components, may be eliminated or reduced through proper experimental design, there will still be plenty of noise sources that obstruct a clean measurement of N400 amplitude. Unstructured noise and ongoing background EEG tend to zero when enough responses are averaged. Averaging across many responses is less of a problem when one wants to study the properties of the N400 itself. The focus of this thesis lies however on deducing the properties of the stimuli, namely the relationship between them. Therefore, averaging across stimuli becomes problematic and other solutions must be found to isolate the N400 from the various noise sources.

Chapter 3 seeks a solution through multivariate filtering. For the purposes of this thesis, a spatial-temporal filter is desired that incorporates knowledge

[4] A look through the collection of publications acquired during the writing of this thesis, 36 studies, even those published as late as 2014, mix a speeded button task with traditional N400 analysis without proper acknowledgement or correction for the overlapping P3b component.

about the shape and timing of the N400 potential, as documented throughout 34 years of N400 literature [7], as well as being able to adapt to the uniqueness of a single recording. We show that by extending the formulation of beamformers, which are classically used in EEG analysis in the context of source localization [108, 122], a flexible spatio-temporal filter can be obtained that can readily make use of such prior information. Because of this, the filter is more flexible in regards to the training data compared with support vector machines (SVMs). For example, the suggested beamformer filter can be designed for a subject, based on training data from other subjects (this is a case of transfer learning), which is demonstrated in chapter 4.

In a simulation study as well as on real EEG data, the performance of the beamformer is compared with a variety of approaches to multivariate filtering. In light of the problematic overlapping P3b component, the techniques are evaluated in terms of ability to isolate the N400 from other ERP components.

1.5.2 Dealing with confound variables

The marriage of a proper experimental paradigm with a good spatio-temporal filtering strategy results in a reasonably good estimate of N400 amplitude. However, this amplitude is modulated not only by the relationship between prime and target, but also a host of properties of the target stimulus (see section 1.3.2).

One way of avoiding confounding effects is to properly counterbalance the data. The baseline N400 amplitude in response to a word can be regarded as an unknown, with the focus instead on relative changes in this amplitude, as the word is presented in different contexts. For example, instead of attempting to quantify the relationship strength of a prime–target word-pair (DOG–TAIL) as an absolute number (34.2), it is easier to determine whether the relation is stronger or weaker compared to another word-pair that combines the target with a different prime (DOG–TAIL < PONY–TAIL).

Where counterbalancing stimuli is unfeasible [123], partial regression and stratification techniques can be used to eliminate confounding variables when they are known [124]. For example, multilevel mixed-effect models [125] can be used to model confounds as random factors and hence perform stratification. Further known confounds can be accommodated by entering them into the model first and continuing the analysis on the residuals, which should be clear of the confound if the assumptions about the nature of the relationship (linear, polynomial, exponential, etc.) were correct.

In order to apply such techniques, any averaging operation across stimuli or subjects should be avoided, which means abandoning the classical ERP technique. The beamformer technique introduced in chapter 3 is capable of estimating N400 amplitude on a single trial basis, clearing the way for advanced statistical models.

1.5.3 Towards the construction of semantic clusters, based on N400 amplitude

In chapter 4 a first attempt is made towards the mapping of semantic structures based only on measurements of N400 component amplitude. Earlier ERP stud-

ies have attempted to classify words into predefined categories (e.g. living versus nonliving [126]). In contrast, chapter 4 attempts to answer the question: "if we assume that this set of words can be neatly organized into two categories, what would those categories be?"

Measurements of the N400 component were performed using the beamformer techniques developed in chapter 3 and were utilized to cluster words together. Words are assigned to a cluster by minimizing the N400 amplitude in response to word-pairs that combine words from the same cluster and simultaneously maximizing the N400 amplitude in response to between-cluster combinations. In this manner, N400 amplitudes are always compared in relation to each-other and never as an absolute, allowing for a proper counterbalancing of the stimuli (see section 4.5).

Chapter 2

Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs

Marijn van Vliet, Nikolay V. Manyakov, Gert Storms, Wim Fias, Jan R. Wiersema, and Marc M. Van Hulle
"Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs"
In: *PLoS ONE* 9.2 (2014), e87650.

2.1 Abstract

This study examines the influence of a button response task on the ERP in a semantic priming experiment. Of particular interest is the N400 component. In many semantic priming studies, subjects are asked to respond to a stimulus as fast and accurately as possible by pressing a button. RT is recorded in parallel with EEG for ERP analysis. In this case, the response occurs in the time window used for ERP analysis and response-related components may overlap with stimulus-locked ones such as the N400. This has led to a recommendation against such a design, although the issue has not been explored in depth. Since studies keep being published that disregard this issue, a more detailed examination of influence of response-related potentials on the ERP is needed. Two experiments were performed in which subjects pressed one of two buttons with their dominant hand in response to word-pairs with varying FAS, indicating a personal judgement of association between the two words. In the first experiment, subjects were instructed to respond as fast and accurately as possible. In the second experiment, subjects delayed their button response to enforce a one second interval between the onset of the target word and the button response. Results show that in the first experiment a P3 component and motor related potentials (MRPs) overlap with the N400 component, which can cause a misinterpretation of the latter. In order to study the N400 component, the button response should be delayed to avoid contamination of the ERP with response-related components.

2.2 Introduction

Semantic priming refers to the case where the presentation of a prime stimulus affects the response to a later target stimulus [3]. When the prime stimulus is related to the target (for example associatively or semantically [127]), the target

is processed more efficiently. An example would be a task where the subject is reading word-pairs, where each word flashes sequentially on a screen. Behavioral responses to the word *dog* will be faster when preceded by the word *cat*, compared to the word *sock*. This increase in efficiency is attributed to our semantic memory [3, 6]. To demonstrate the priming effect one can measure the RT on a task that requires the subject to process the stimuli [12]. It can also be measured with electroencephalography (EEG), where it manifests as an event-related potential (ERP) component called the N400 [7, 75].

Many semantic priming studies use a task that requires the subject to look for some property of the stimulus currently presented [3, 14]. For example, a commonly used task is lexical decision [8], in which the subject is asked to make a decision about whether the presented string of letters is a valid word or a non-word. The subject presses one of two buttons as quickly as possible to give a response. The difference in RT is analyzed between valid word strings that were preceded by a related stimulus and valid word strings that were preceded by an unrelated one. Since such a task is known to generate a P3 component of which the latency correlates with the RT rather than the onset of the target stimulus [120, 121], we classify the component as response-related as opposed to stimulus-locked. Response-related components can be visualized by cutting EEG segments locked to response onsets and average them to get a response-locked ERP [128].

When conducting a semantic priming experiment that is designed to study the N400 component, Kutas et al. [74, 76], Duncan et al. [49] and Picton et al. [14] discourage the use of a button press in the time window used to analyze the stimulus-locked ERP, because a P3 component may be generated that overlaps with the N400 [15]. This recommendation is made as a side note, but deserves more attention as studies, that use a task where the subject presses a button during the ERP time window, continue to be published. For example, studying the method sections of the results of a search using Scirus[1], with the query `"N400" "reaction time" "response time"` where the results were limited to journal articles published in 2011 alone, yielded seven semantic studies that mixed recording RT on a button press with ERP analysis. Two of them analyze both the stimulus-locked and response-locked ERPs, while the rest disregard response-related effects on the ERP completely.

Because of this, we feel a study dedicated to the problem is in order to examine the generated components in more detail. In this study, we try to gauge the severity of the distortion and the implications for the resulting conclusions drawn from such data. For this purpose, two experiments were performed: one where the subjects performed a speeded button response task and one where the response was delayed. The generated ERPs are analyzed to demonstrate the risk of contaminating stimulus-locked potentials, such as the N400, with response-related ones, such as the previously mentioned P3.

[1] www.scirus.com

2.3 Materials and Methods

In a semantic priming study across two experiments, subjects read a series of sequentially presented words, organized in pairs. In the first experiment the subjects pressed one of two buttons to indicate whether the two words of a word-pair were related or not as quickly as possible while remaining accurate in their decision making. In the second experiment the subjects performed the same task but delayed their button response until a cue was given. The two experiments will be referred to as the *speeded condition* and *delayed condition* respectively.

2.3.1 Subjects

The experiment employing the speeded condition was performed with 10 university students (3 female, aged 19–27 years), all right-handed and native speakers of Flemish-Dutch. As the main interest of this study is the effect of delaying the subject's button response, the experiment employing the delayed condition was performed with the same subjects from the first experiment to reduce between-subject variability. Because the recordings of the speeded condition were already completed before the inception of this study and the construction of the delayed setting, all subjects performed the speeded task first, followed by the delayed task at a later time. Since memory effects influence the N400 potential [129], subjects performed the latter experiment a minimum of 2 months after the first to mitigate these effects [101].

2.3.2 Ethics Statement

This study was approved by the UZ Leuven ethics committee. All subjects were volunteers and signed an informed consent form before each experiment.

2.3.3 Materials

To construct word-pairs with varying associative relatedness, we used the association norm dataset, compiled by De Deyne and Storms [46, 47].

The stimulus list used during the experiments consists of a total of 800 Flemish-Dutch word-pairs, selected (figure 2.1) with varying FAS from the association norm dataset mentioned above. FAS was determined through a free association task, where cue words were presented to 100 subjects. They wrote down the first three words that came to mind to each cue [46, 47]. The FAS of a (prime, target) word-pair is defined as the number of subjects that wrote down the target word in response to the prime word. In this study, only the first association of each subject is considered. The stimulus list consists of the top 100 strongest related word-pairs (FAS ranged 69–95, mean FAS = 75.62) and 100 word-pairs where the prime and target words were randomly chosen and no record of the word-pair existed in the association norm data, therefore having an assumed FAS of 0. The remaining 600 word-pairs were chosen such that the logarithm of their FAS score is uniformly distributed on the range $[0 \ldots 69]$, extending the

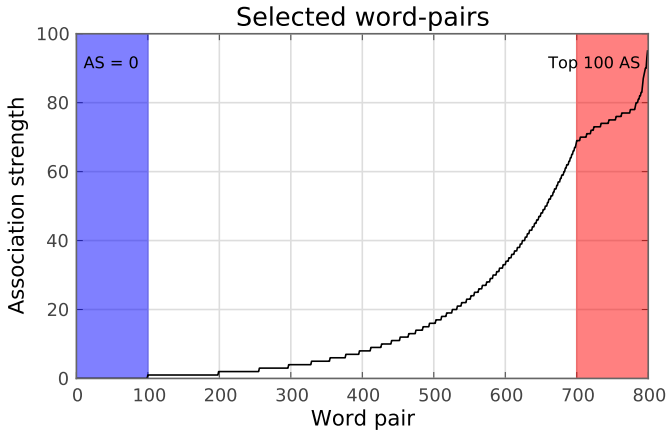


Figure 2.1: FAS of all 800 word-pairs that were used. The blue shaded area contains 100 word-pairs with an FAS of zero, meaning the words are completely unrelated. The red shaded area contains the 100 word-pairs with the highest FAS in the association norm data. Both words are four to six letters in length. The FAS of the remaining word-pairs follows a log scale.

complete range between the unrelated and the top 100 strongest related word-pairs. The log scale was chosen because when the association norm data were analyzed, some properties of the word-pairs that co-vary with the FAS, correlate better with its logarithm than the raw values. For example, the length of the target word ($\rho = -.18$ and $\rho = -.12$ respectively) and the in-degree of the word-pair ($\rho = .17$ and $\rho = .13$ respectively). A word's in-degree is the number of unique words to which the participants in the free association study generated the target word. This is a measure of the centrality of the word if the norm dataset is visualized as a semantic network [46]. Based on these logarithmic relationships, we hypothesize that the relationship between RT/N400 and FAS might also be logarithmic. This hypothesis is tested in the results section.

All selected words for the stimulus list have a length of 4 to 6 letters and only reasonably common words were chosen. This was achieved by limiting the minimum word frequency to 2 occurrences per 10^6 words in the SUBTLEX-NL corpus [130]. Also, the second word of a word-pair (i.e. the target word) had a minimum in-degree of 5, meaning they were well connected in the association norm data, which is an indication that they should be familiar.

In addition to capturing the button response of the participant, EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System (BioSemi, Amsterdam, the Netherlands), having a 5th order frequency filter with a pass band from 0.16 Hz to 100 Hz, and sampled at 2048 Hz. An electrooculogram (EOG) was recorded as recommended by Croft et al. [131]. Two electrodes were placed on both mastoids and their average was used as a reference for the EEG.

Stimulus presentation was done using MATLAB (MathWorks, Natick, Massachusetts, U.S.A.) and Psychtoolbox [132]. EEG data was processed using the NumPy/SciPy Python packages [133]. Figures were created using the Matplotlib Python package [134]. Linear mixed effects (LME) models were fitted using the LME4 package for R [125].

2.3.4 Experimental procedure

Subjects were seated in an upright position approximately one meter from a computer screen. The dominant hand rested upon the table with the index and middle fingers resting on mouse buttons.

A trial consisted of the sequential presentation of a single word-pair. The stimuli were shown as white text on a black background in the Arial font with a point size of 50 and centered on the screen both horizontally and vertically. The subject was instructed to press the left mouse button to indicate the prime and target were related or the right mouse button to indicate they were not. This can be seen as a simplified version of the JAM task [17]. Responses were performed by the index and middle fingers of the dominant hand, which was the right for all subjects. The mapping of the response to the mouse buttons and the hand used for responding were not counterbalanced. Normally, the hand used to respond and the mapping of the mouse buttons to *related* and *unrelated* responses is counterbalanced across recordings. This reduces differences due to left/right lateralized effects in the grand average RTs and ERPs. This study however analyzes potentials generated by the response of the subjects by averaging response-related components across subjects. Therefore, we tried to reduce the variability between the responses as much as possible by making all subjects respond with their dominant hand (for all subjects, the right hand) at all times and did not counterbalance the mapping of the mouse buttons. It is possible that due to the lack of counterbalancing, the spatial location of ERP components such as the N400 and P3 influenced by this fixed response button assignment.

In each experiment, 20 trials using word-pairs that are not part of the stimulus list, were presented for the subject to practice. Next, 800 trials, split up into 5 blocks of 160, were presented. Between each block the subject was prompted to take a short break. Each experiment lasted between 35 and 45 minutes, depending on the length of the breaks.

In the speeded condition, during each trial, the prime word was presented for 200 ms and the target-word for 1000 ms with a SOA of 500 ms. Between trials, a blank screen was presented for one second. In the speeded condition, the subjects were told to respond as quickly and accurately as possible upon being presented with the target word. The delayed condition employed the same procedure with the exception that the target word would stay on the screen for 1500 ms, turning from white to yellow after 1000 ms. Subjects were told to delay their response until after the target word had changed color. In both conditions, the subjects had 1000 ms to respond or a no-response code would be logged instead.

2.3.5 Data preprocessing and extracting trials

The EEG was bandpass filtered offline between 0.1 Hz to 50 Hz by a 3rd order two-way infinite impulse response (IIR) filter to attenuate large drifts and irrelevant high frequency noise, but retain eye movement artifacts. It was downsampled to 256 Hz afterwards. The EOG was used to attenuate eye artifacts from the EEG signal using the regression method outlined in [131]. As we are mostly interested in examining N400, P3 and motor related (MR) components, the EEG signal was band pass filtered again, from 0.5 Hz to 15 Hz by a 3rd order two-way IIR filter, to further attenuate signals that are not of interest in this study. After frequency filtering, the signal was cut into segments from 10 ms before the onset of the target stimulus to 1500 ms after. Baseline correction was performed using the average voltage in the 10 ms interval before the stimulus onset as baseline value. Finally, trials in which the subject had not made a button response, or was too late in making a response, were discarded.

2.3.6 Statistical analysis

Statistical analysis of various effects were done by means of an LME model. For all usages of LME models in this article, random effects consisted of subjects (modeling both slopes and intercepts) and word-pairs (modeling intercepts only). Models were fitted using restricted maximum likelihood (REML). This design follows the recommendations of Baayen et al. [135]. Because the degrees of freedom in an LME model are non-trivial, degrees of freedom and the resulting p-values were estimated using Satterthwaite's approximation [136].

When testing for a correlation between two vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}$ was entered as the dependent variable in the LME model and $\mathbf{y}$ as a fixed effect (subjects and word-pairs were random effects). When testing for a difference between the means of two groups $\mathbf{x_1}$ and $\mathbf{x_2}$, we concatenated the values into a single vector $\mathbf{x}$ and used a coding vector $\mathbf{y}$ to label each value $x \in \mathbf{x}$ with a corresponding $y \in \mathbf{y}$, where $y = 0$ if $x \in \mathbf{x_1}$ and $y = 1$ if $x \in \mathbf{x_2}$. In the LME model, $\mathbf{x}$ was entered as the dependent variable and $\mathbf{y}$ as a fixed effect (subjects and word-pairs were random effects). In both cases, we report the obtained regression weight (w), the t-value (t) and p-value (p).

2.4 Results

For both conditions 8000 trials (10 subjects × 800 trials) were initially collected. Rejection of no-response trials brought this number down to 7759 (3.0% rejected) EEG sweeps for the speeded and 7949 (0.6% rejected) for the delayed condition. For each subject, ERPs were constructed by sorting the trials by either the FAS of the stimulus or the RT, grouping them into 8 non-overlapping bins of equal size (table 2.1) and averaging the trials in eacencephaloh. Finally, the ERP of each bin was averaged across subjects to form the grand-average stimulus-locked ERPs and response-locked ERPs.

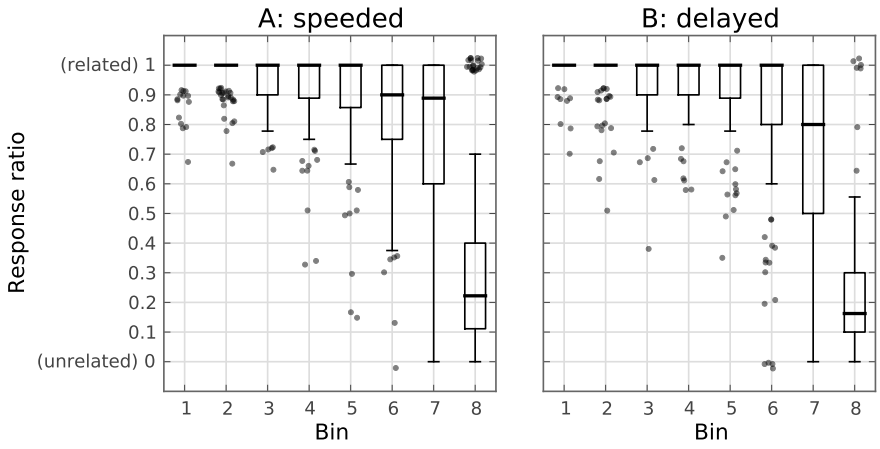


Figure 2.2: Boxplots of button responses made by the subjects. Each boxplot corresponds to a bin (see table 2.1). Whiskers extend to the inner quartile range. Outliers are plotted as semi-transparent dots, with some random jitter applied to the x (±0.4) and y (±0.05) position to reduce overlap. Bin 1 contains trials with a high FAS and subsequent bins contain trials with lower FAS. Bin 8 contains mostly trials with FAS= 0 (the trials in this bin with FAS> 0 are mostly outliers in the corresponding boxplot). The y-axis shows the number of subjects that rated the two words of the word-pair as being related, divided by the total number of subjects (10). **A**: responses during the speeded condition. **B**: responses during the delayed condition.

2.4.1 Button responses

In the speeded condition, RT shows an inverse dependency to the FAS of the word-pair (table 2.1). Statistical analysis of the effect was done by means of an LME model with RT as the dependent variable and FAS as a fixed effect (subjects and word-pairs were random effects). Two models were constructed: one that used the raw FAS values and another that used the logarithm of the FAS (dropping the trials in which FAS was zero). The model using the log FAS provided a better fit on the data (log likelihood -10098) than the model using the raw FAS values (log likelihood -10166). For the speeded condition, the model indicated a significant effect of log FAS on RT ($w = -0.0153$, $t = -6.95$, $p < 0.0001$). No effect was found for the delayed condition ($w = -0.000379$, $t = -0.121$, $p = 0.906$). It is likely the subject had already prepared his/her decision whether the two stimuli are related in advance.

During the experiment, each word-pair was rated by the 10 subjects by either pressing the left (unrelated) or right (related) mouse button. We must point out that the purpose of the response task was to keep the subjects focused during the experiment, rather than obtaining reliable JAM ratings. No corrections were for example performed for response bias due to the order of the stimuli (after a long series of unrelated word-pairs, a subject would be biased toward rating a new word-pair as related). We refer to the number of times a "related" JAM response was given, divided by the total number of responses, as the *re-*

sponse ratio. Table 2.1 shows the response ratio for each bin and it can be seen that as the FAS between words increases, the likelihood of a "related" JAM response increases as well. A break from the overall trend occurs between word-pairs with FAS≈ 0 (bin 8, sorted by FAS) and FAS≈ 1 (bin 7, sorted by FAS) as the response ratio drops sharply. This contributes some evidence that a log scale is suitable for FAS, as $\log 0 = -\infty$ and $\log 1 = 0$. figure 2.2 shows the distribution of the response ratios of the word-pairs in each bin, where the response ratio corresponds to the number of subjects that gave a "related" JAM response to the word-pair, divided by the number of subjects (10). These response ratios show a pattern, which is similar during both the speeded and the delayed conditions 2.2. At low FAS levels, instead of all subjects agreeing that the words are unrelated, a high variance is seen. Indeed, a good portion of the word-pairs with FAS=1 were unanimously rated as related by all subjects. As also shown in a study by Maki [17], subject's JAM ratings are generally higher than the free association scores.

2.4.2 ERPs during the speeded condition

The ERPs recorded during the speeded condition suggest a strong N400, which becomes more negative as the FAS between the words becomes smaller (figure 2.3A). The timing and scalp topography of this effect is very similar to the one described in the literature [7, Fig. 1]. Statistical analysis was performed on the difference between the first and last bins. For each individual EEG sweep belonging to either the first or the last bin, the voltage at electrode Pz over the time-range 300 ms to 500 ms was used to quantify the candidate N400 component. An LME model was constructed to test the difference of the average EEG voltage between the first and last bins (see the methods section for details), which was found to be significant ($w = -5.22$, $t = -5.44$, $p < 0.0001$).

When comparing short versus long RTs, it becomes clear that more processes are going on in the same time window, as a large component is now seen aligned to the mean RT of the bin (figure 2.3B). A statistical analysis of the latency of this component, using a template matching technique, is given in a later section. This component is also present in the response-locked ERPs (figure 2.4, thick line). The topographies of the different components are very similar: posteriorly and slightly to the left. The latter most likely due to the fact that all subjects responded using their right (for all subject, the dominant) hand.

A similar component was described in various RT experiments [120, 137], and was in those studies classified as a P3. Kutas et al. (1977) observed that during a speeded task, the P3-latency and RT are strongly correlated, but not necessarily equal. This lead to the conclusion that the P3-latency could be a correlate of stimulus processing time. Since FAS and RT are also correlated (as shown in the previous subsection), bins with a different mean FAS will also have a different mean RT (figure 2.3A, vertical lines). It is likely that the effect seen in the FAS-binned case consists of not only the N400, but is in fact dominated by the difference in latency of the P3 component seen in the RT-binned case, which has a similar scalp topography and overlaps in time with the N400.

2.4.3 ERPs during the delayed condition

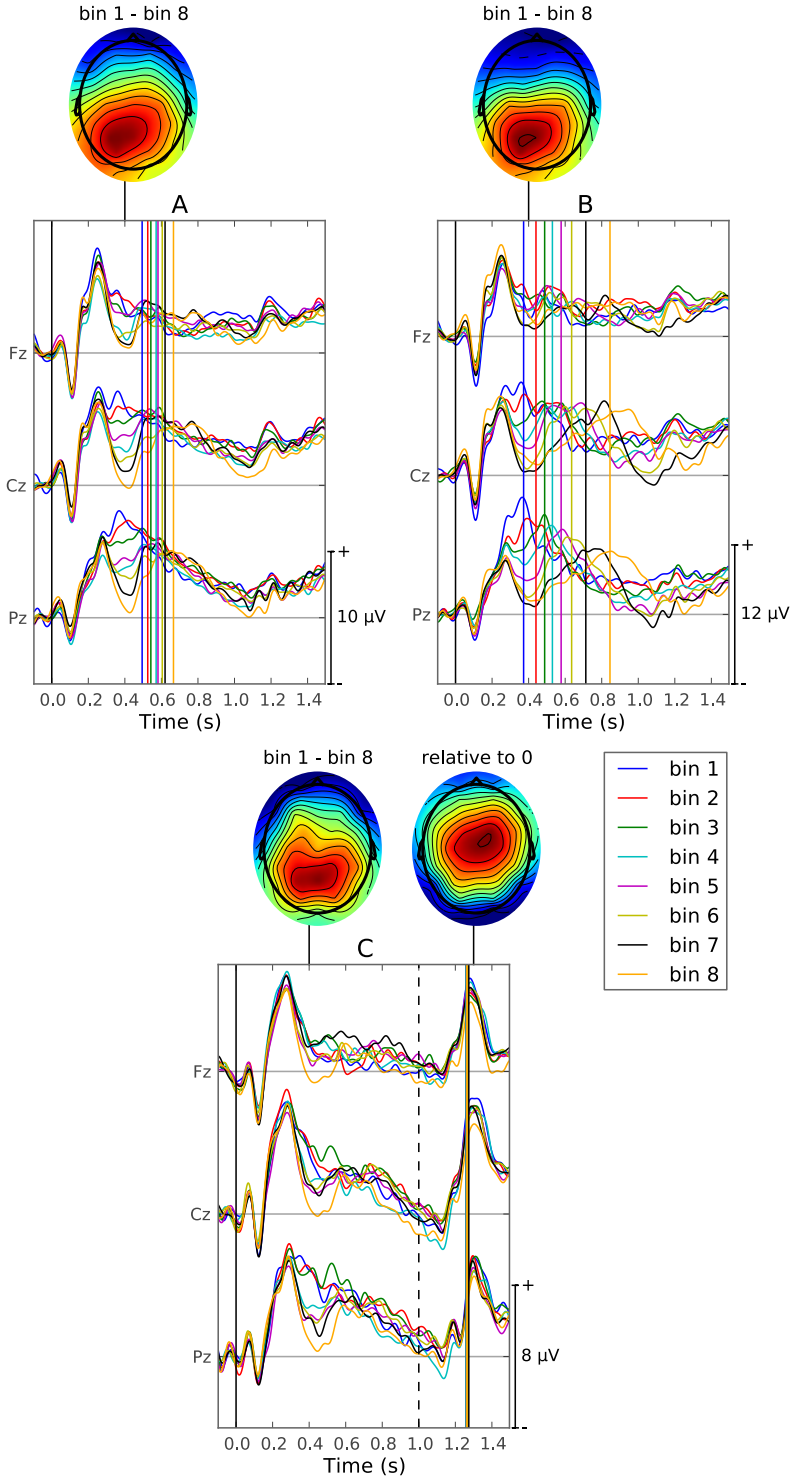
During the delayed condition (figure 2.3C), a component is seen which is very similar to the one described in the speeded condition. Namely it occurs at 400 ms after the onset of the target word and is increasingly negative as the FAS of the bin decreases. Analyzing the difference in mean EEG voltage of electrode Pz in the time-range 300 ms to 500 ms, between the first and last bins, shows a significant difference ($w = -3.13$, $t = -2.46$, $p = 0.0276$). Additionally, a strong component can be seen at 1300 ms, 300 ms after the target word turned yellow, which cued the button response of the subject. Although the mean latency of this component is close to the mean RT (figure 2.3C, vertical lines), it is not visible in the response-locked ERP (figure 2.4, thin line). Judging from the timing and scalp topography of this component, it's likely a P2 generated by the response cue [54]. Similar P2s can be seen around 300 ms after the onset of the word stimulus during both the speeded and delayed conditions. The response-locked potentials for the delayed condition contain mostly MRP with a similar shape as described in the literature [138, 139]: a negative slope (NS) leading up to a negative, mostly anterior, MRP at the moment of the button press (figure 2.4, thin line).

The components generated during the delayed condition (figure 2.3C and figure 2.4, thin line) are spatially more clearly separated. The N400 occurs central-posterior, whereas the P2 component at 1300 ms occurs slightly anteriorly, slightly to the right. The MRP (figure 2.4, delayed condition) displays a dipole pattern at the onset of the button press with the positive component at a left-posterior location and the negative component central-frontal.

2.4.4 Analyzing P3-latencies

To further study the P3 component observed in the speeded task, we repeated the analysis done by Kutas et al. [120] and attempted to estimate its latency for each trial. For each subject, a slightly modified version of the template matching technique developed by C. D. Woody [140] was applied. Since this technique operates on one-dimensional data, we limited the analysis to the Pz electrode:

Figure 2.3 ***(facing page)*****: Stimulus-locked ERPs and scalp topographies for both experimental conditions.** Vertical lines are plotted to show the mean RT of the trials belonging to each bin (table 2.1), using the same color code as the ERPs. Scalp topographies are drawn for components of interest, using red for positive and blue for negative values. Note that the y-axes have different scales and each scalp topography uses its own normalized scale. **A**: Stimulus-locked ERP of the speeded condition; trials sorted by decreasing FAS. The scalp topography shows the difference between bins 1 and 8 at 400 ms. **B**: Stimulus-locked ERP of the speeded condition; trials sorted by increasing RT. Scalp topography shows the difference between bins 1 and 8 at 400 ms. **C**: Stimulus-locked ERP of the delayed condition; trials sorted by decreasing FAS. A vertical dotted line indicates the moment the target word turned yellow, which cued the response of the subject. Two scalp topographies are drawn, one showing the difference between bins 1 and 8 at 400 ms, the other showing the voltage at 1300 ms relative to zero.


bin 1 - bin 8
A
Fz
Cz
Pz
10 µV
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Time (s)
bin 1 - bin 8
B
Fz
Cz
Pz
12 µV
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Time (s)
bin 1 - bin 8
relative to 0
C
Fz
Cz
Pz
8 µV
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Time (s)
bin 1
bin 2
bin 3
bin 4
bin 5
bin 6
bin 7
bin 8

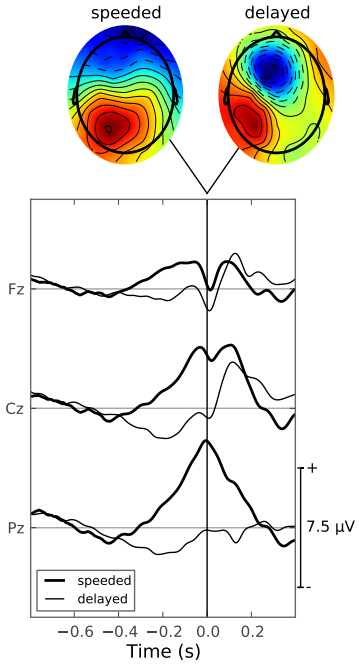


Figure 2.4: Response-locked ERPs and scalp topographies for both experimental conditions. Shown are the grand response-locked ERPs across all bins and all subjects for the speeded condition (thick line) and delayed condition (thin line). Scalp topographies are given showing the voltage at 0 ms, using red for positive and blue for negative values. Note that both scalp topographies use their own normalized scale.

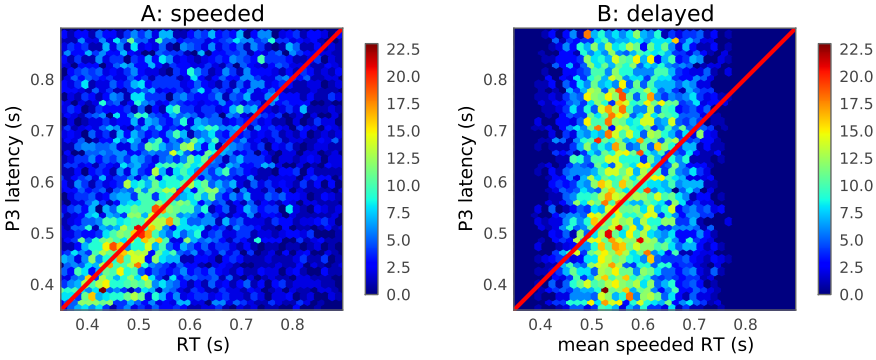


Figure 2.5: Two-dimensional histogram of single trial component latency versus RT. A hexagonal grid shows histograms of component latency versus RT for all non-rejected trials of all subjects. The values along the x and y axes indicate seconds since the onset of the target stimulus. Component latency was determined through iterative template matching (see the results section). **A**: P3 latency during the speeded condition; superimposed as a red line is $x = y$. **B**: P3 latency during the delayed condition, versus the mean RT of each word-pair collected during the speeded condition.

1. Filter the signal with a 3rd order, two-way, Butterworth filter between 0.5 and 6 Hz in order to remove alpha activity.

2. Take the RT for each trial as the initial P3-latency estimate.

3. Based on the P3-latency estimates, cut the EEG signal for each trial between −500 ms to 500 ms, relative to the latency estimate. Average the cuts to generate a P3-locked ERP. This ERP becomes the new template.

4. For each trial: calculate the cross-correlation of the signal with the template. We restricted this analysis to the time window from 350 to 900 ms after the onset of the target stimulus. This yields a vector containing for every sample a score. Determine the peaks in the score vector by detecting sign changes in the derivative. Take the position of the largest positive peak to be the new estimated P3-latency.

5. Repeat steps 3–4 four times to refine the template.

For the speeded condition, it can be seen that, on average, the P3-latency follows the RT (figure 2.5A), explaining the alignment of the two in figure 2.3B. However, the P3 component is not strictly aligned on the response, sometimes occuring before or after the onset of the button press. This finding is consistent with Kutas et al. [120]. They postulated that the P3 is an index of the time it takes to process the stimulus and make a decision. Statistical analysis of the effect was done by using P3-latency as the dependent variable and RT as the fixed effect of the LME model. For the speeded condition, the model indicated a large effect of RT on P3-latency ($w = 0.216$, $t = 14.4$, $p < 0.0001$). Repeating the analysis with the logarithm of FAS as fixed variable also yielded a significant effect ($w = -0.00736$, $t = -5.82$, $p < 0.0001$), which comes as no surprise as the RT was shown above to correlate strongly with FAS. However, the model using RT as a fixed effect has a much better fit to the P3-latency data (log likelyhood 3838) than the model using FAS (log likelyhood 3766). Presumably, this is because semantic priming is influenced by many factors besides FAS, such as word length and frequency. These factors are all reflected in the RT of the subject.

There is a possibility that a P3 is also generated in the delayed condition during the time leading up to the response cue as the subject makes up his/her mind. We employed the template matching technique to find P3-latencies in the 350 to 900 ms window after the onset of the target stimulus. However, since in the delayed condition, RTs cannot be used as the initial P3-latency estimate in step 2, since RTs generated during this condition are all outside of the 350 ms to 900 ms time window. Instead, the final template used to estimate the P3-latency during the speeded condition was re-used as initial template for the delayed condition. The obtained P3-latencies were analyzed with an LME model. No significant effect was found using FAS as a fixed effect ($w = -0.00251$, $t = -1.68$, $p = 0.119$). We observed earlier that, when using a speeded task, RT was a better predictor of P3-latency than FAS. For each word-pair, we calculated the mean RT, obtained during the speeded condition, and used it as a predictor for the P3-latency during the delayed condition (figure 2.5B). This "mean speeded RT"

predictor variable does show a very small, but significant effect when used as fixed effect in the statistical model ($w = 0.133$, $t = 4.58$, $p < 0.0001$), despite being hardly noticeable in figure 2.5B. This still leaves open the question whether during the delayed condition, a P3 is generated which latency correlates with the FAS of the word-pair, since the effect is too small to be reliably detected by our template matching technique, which was successful in detecting the P3 generated during the speeded condition.

2.4.5 Impact of P3-latency on overall ERP

We noted before that the P3 potential observed in the speeded condition overlaps in time and scalp topography with the N400 as described in the literature. To demonstrate how a difference in P3-latency can mask N400 effects, we made four groups of trials in such a way that N400 and P3 effects are in competition. The first group consists of trials recorded during the speeded condition, where the stimulus FAS was low (≤ 1) and the P3-latency was short (< 600 ms). The second group consists of trials also recorded during the speeded condition, where the stimulus FAS was high (> 20) and the P3-latency was long (> 700 ms). Note that we would expect the first group to have a more negative N400 than the second group (based on FAS), but also a shorter P3-latency (based on the latency estimates). Equal group sizes were enforced by randomly discarding trials from the larger group. The third and fourth groups consisted of the trials recorded during the delayed condition that correspond in terms of subject and stimulus to the trials in the first and second groups. In case a subject-stimulus combination was not available due to rejection of no-response trials, it was removed from all groups. In the end, all groups contained 647 trials. The P3-latency of the chosen trials are unusual, because normally a trial with a high AS stimulus would result in a short P3-latency. The response ratios for the four groups are 0.52, 0.96, 0.96 and 0.96 respectively. The FAS of the trials in groups 1 and 3 would place them in bins 7–8 (sorted by FAS) in table 2.1, which have an average response ratio of $\frac{.72+.26}{2} = .49$ in the speeded and $\frac{.71+.21}{2} = .46$ in the delayed condition. The FAS of the trials in groups 2 and 4 would place them in bins 1–3, which have an average response ratio of $\frac{.98+.97+94}{3} = .96$ in the speeded and $\frac{.98+.97+.95}{3} = .97$ in the delayed condition. The response ratios of the trials in the four groups are representative of those of the entire dataset, even if the P3-latencies are unusual.

Comparing the ERPs of the four groups shows that the P3 component dominates the ERP in the speeded condition (figure 2.6A). While the low FAS group portrays a distinctive negative peak at 400 ms, the P3 occurs shortly afterwards, causing a net positive difference with the high FAS group. In other terms, the P3-latency effect "wins" over the N400 effect. Analyzing the mean voltage between 400 and 500 ms for electrode Pz, yields a significant difference between the two groups ($w = -3.35$, $t = -4.87$, $p = 0.000185$). This would us to to conclude that the target-words of the first group were processed faster than the target-words of the second. However, we must be careful not to attribute this to a property inherent to the word-pairs, such as declaring that for these word-pairs the sub-

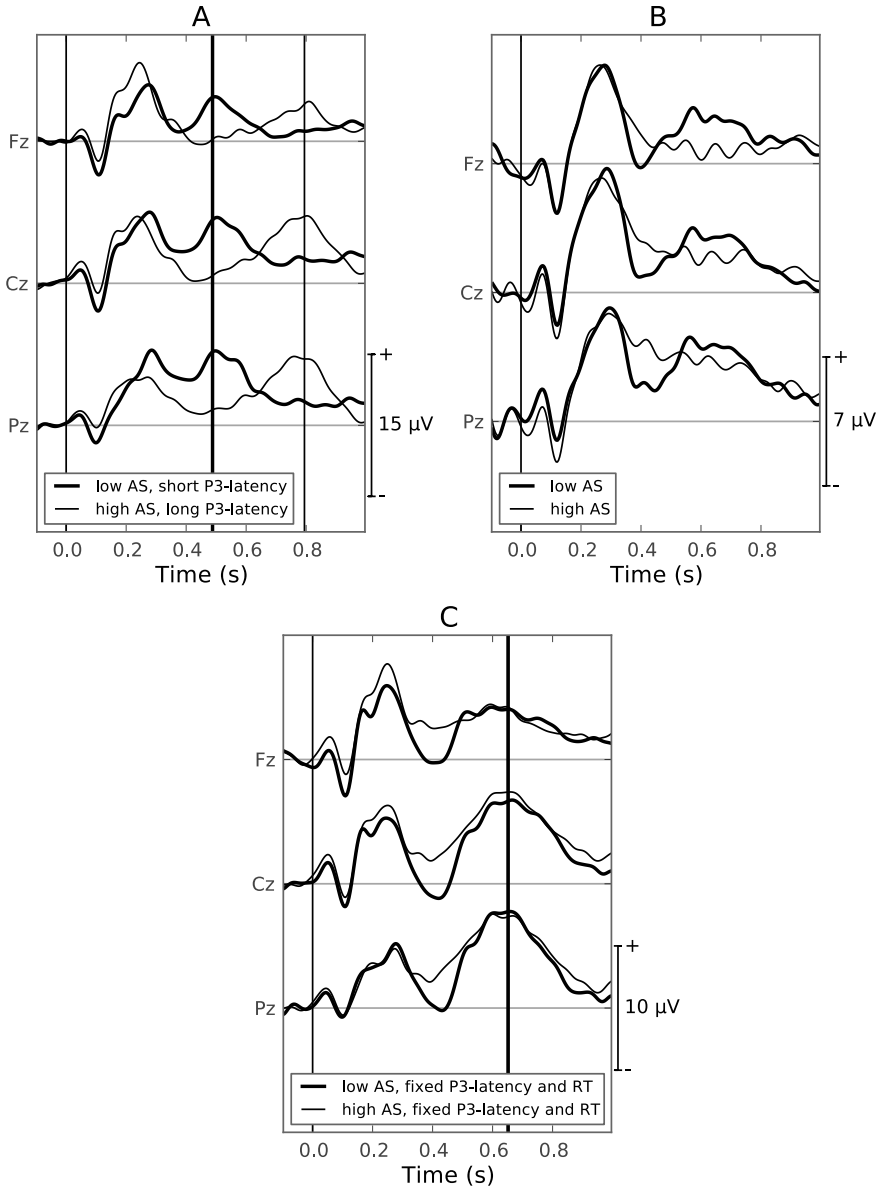


Figure 2.6: Grand average ERPs demonstrating the effect of differences in P3-latency. **A**: Speeded condition, low FAS (≤ 1) and short P3-latency (≤ 600 ms, thick line) versus high FAS (> 20) and long P3-latency (> 700 ms, thin line). Vertical lines indicate the mean P3-latency of each group. **B**: Delayed condition, using the same subject/word-pair combinations used to create A. **C**: Speeded condition, low FAS (≤ 1, thick line) versus high FAS (> 20, thin line) using a fixed P3-latency and RT (around 650 ms). Vertical lines indicate the mean P3-latency of each group, which overlap in this case.

	Speeded condition, sorted by FAS			Speeded condition, sorted by RT		
bin	mean FAS (std)	mean RT (std)	resp	mean FAS (std)	mean RT (std)	resp
1	76.05 (5.27)	0.49 (0.122)	0.98	35.86 (30.49)	0.37 (0.035)	0.99
2	49.51 (9.98)	0.53 (0.127)	0.97	30.85 (28.31)	0.44 (0.015)	0.99
3	24.50 (5.03)	0.54 (0.131)	0.94	27.01 (27.27)	0.49 (0.013)	0.96
4	11.95 (2.52)	0.57 (0.138)	0.92	22.41 (25.74)	0.53 (0.013)	0.91
5	05.66 (1.29)	0.58 (0.139)	0.90	17.94 (23.42)	0.58 (0.015)	0.82
6	02.57 (0.63)	0.60 (0.148)	0.82	13.76 (20.93)	0.64 (0.019)	0.68
7	01.06 (0.23)	0.62 (0.146)	0.72	13.65 (21.96)	0.71 (0.026)	0.61
8	00.03 (0.17)	0.67 (0.145)	0.26	09.69 (17.40)	0.85 (0.059)	0.54
	Delayed condition, sorted by FAS					
bin	mean FAS (std)	mean RT (std)	resp			
1	75.93 (5.29)	1.26 (0.131)	0.98			
2	48.60 (10.01)	1.27 (0.131)	0.97			
3	23.76 (4.93)	1.26 (0.129)	0.95			
4	11.52 (2.44)	1.26 (0.133)	0.94			
5	05.46 (1.22)	1.26 (0.132)	0.90			
6	02.48 (0.57)	1.27 (0.130)	0.85			
7	01.02 (0.15)	1.27 (0.144)	0.71			
8	00.01 (0.11)	1.26 (0.134)	0.21			

Table 2.1: Descriptive statistics of the bins across all conditions. For each condition, the mean and standard deviation of FAS and RT are listed, as well as the response ratio of each bin. The response ratio is the number of "related" JAM responses given by the subjects, divided by the total number of responses. The conditions correspond to figure 2.3A, 2.3B and 2.3C respectively. Units for RT are seconds.

jects disagree with the association norm data. When we look at the delayed condition (figure 2.6B), using matched subjects and matched stimuli, the N400 component is no longer visibly obscured by the P3. Based on the N400 amplitudes of both groups, we would now draw the opposite conclusion in this case (comparing the mean voltages between 400 and 500 ms of electrode Pz in a similar fashion as before: $w = 2.84$, $t = 1.52$, $p = 0.157$).

Finally, two more groups of trials were created using trials recorded during the speeded condition: a low-FAS (≤ 1) and a high-FAS (> 20) group, where all trials have a similar P3-latency and RT. This should let us compare the difference in N400 amplitude between the low-FAS and high-FAS cases, without distortion due to differences in P3-latency. Consider the following scoring function:

$$s = |\hat{t}_{\mathrm{RT}} - t_{\mathrm{RT}}| + |\hat{t}_{\mathrm{P3}} - t_{\mathrm{P3}}|,$$

where t_{RT} is the RT of the trial and t_{P3} is the estimated P3-latency of the trial. Low values of s correspond to trials with RT close to $\hat{t}_{\mathrm{RT}}$ and P3-latency close to $\hat{t}_{\mathrm{P3}}$. To construct the low-FAS group, for each subject, out of all trials with FAS≤ 1, the 75 trials with the lowest score s were selected. In the scoring function, both $\hat{t}_{\mathrm{RT}}$ and $\hat{t}_{\mathrm{P3}}$ were set to 650 ms to avoid overlap of the P3 component with the N400. The high-FAS group consists of trials that correspond to the trials in the low-FAS group in terms of subject, RT and estimated P3-latency. For each trial in the low-FAS group, the scoring function, with $\hat{t}_{\mathrm{RT}}$ set to the RT of the low-FAS trial and $\hat{t}_{\mathrm{P3}}$ set to its P3, was used to score all trials of the same subject and with FAS> 20. The trial with the lowest score was selected. The end result were two

groups of 750 trials with an equal mean P3-latency ($w = -0.000294$, $t = -0.0367$, $p = 0.971$) and mean RT ($w = -0.0268$, $t = -1.85$, $p = 0.094$). Comparing the ERPs of both groups (figure 2.6C) shows the N400 potential during the speeded condition without interference of the P3 (comparing the mean voltages between 400 and 500 ms of electrode Pz: $w = 2.95$, $t = 2.84$, $p = 0.0143$).

2.5 Discussion

During the experiments, the subjects were asked to read a word-pair, rate it either as related or unrelated and press the corresponding button. Based on the stimulus- and response-locked ERPs, we can identify at least four sources of ERP components. When a stimulus is displayed on the screen, a series of components, among which a strong P2 component, is evoked by it. Next, the experiments were designed to evoke a priming effect that is known to generate an N400, so this component is expected to be present in both speeded and delayed conditions. This component is linked to the semantic processing of the words [7]. Additionally, the subject is required to choose which button to press. In the speeded condition this occurs as soon as the subject has decided whether the words are related and the decision generated a P3 component close to the moment when the button was pressed. This causes the P3 to be best visible in the response-locked ERP. In the delayed condition, during the one second interval between the onset of the target word and the onset of the response cue, the subject has time to decide whether the word-pair is related or not, so one might expect a P3 component to occur. However, the template matching technique that was successful in showing a positive relationship between P3-latency and FAS during the speeded condition failed to do so during the delayed condition. Finally, when the subject presses the button, response-locked MRPs are generated leading up to, as well as occurring at, the onset of the button press [139].

In the speeded condition, the N400, P3 and MRP components have overlapping time windows, leading to a mixture that presents itself in the stimulus-locked ERP, in the low-FAS versus the high-FAS case, as a difference wave very similar to the N400 component alone (figure 2.3A). Even if the N400 would not be generated at all, the combination of the P3 and MRP added together forms a mixture (figure 2.4, thick line) which overlaps in time and topography with that of the N400 component. This makes it very difficult to accurately assess the magnitude of the N400 present in the EEG signal, leading to conclusions that are more likely based on the P3-latency than the N400, as demonstrated in figure 2.6A-B. To examine N400 effects alone, we demonstrated a template matching technique that can be employed to detect single trial P3-latencies, allowing a researcher to compare groups of trials, keeping both P3-latency and RT fixed (figure 2.6C).

2.6 Conclusion

If the goal of the experiment is to capture N400 effects, we advise caution when the subjects perform a button response close to the time window of interest for

ERP analysis. We presented evidence that large P3 and MRP components overlap with the N400, which causes difficulties isolating the latter. These findings justify the advise of Picton et al. [14], Kutas et al. [74, 76] and Duncan et al. [49] to not employ a response task in the same time window used when analyzing the N400. Where they merely advise against it without elaborating on the subject, our study demonstrates the severity of the issue. To study the N400 it is recommended that the subject is given an explicit task to keep him/her alert [14]. We recommend a design where the button response is delayed to avoid contamination of the ERP with response-locked components.

When a study requires RT data, and RT data cannot be acquired during a separate recording, one has to deal with the P3 and MRP components in some way. We demonstrated a simple template matching technique, developed by Woody [140], to estimate single trial P3-latencies. Afterwards, N400 effects can be gauged by comparing groups of trials with equal mean P3-latency and RT. We encourage the reader to also look into spatial decomposition techniques such as ICA, which was employed successfully by Jung et al. [141], and temporal decomposition techniques, such as the one proposed by Takeda et al. [128], to separate response-locked and stimulus-locked components. A thorough discussion of these techniques is beyond the scope of this paper.

The argument can be made that the main conclusions of many linguistic studies are about whether a priming effect occurs in a certain condition or not. In this case it does not matter what components dominate the ERP, as long as a priming effect is demonstrated. While this reasoning is correct, we counter that in this case one can suffice with reaction time recordings only. Often the purpose of jointly recording reaction time and EEG is to gather *additional* information about the semantic processes in our brains. In this case one must be aware of the different components that are in play and not for example mistake a difference in P3-latency for an N400 component.

Chapter 3

Single-trial ERP component analysis using a spatio-temporal LCMV beamformer

Marijn van Vliet, Nikolay Chumerin, Simon De Deyne, Jan Roel Wiersema,
Wim Fias, Gerrit Storms and Marc M. Van Hulle
"Single-trial ERP component analysis using a spatio-temporal LCMV beamformer"
In: *IEEE Transactions on Biomedical Engineering*, (in press).

3.1 Abstract

Goal: For statistical analysis of ERPs, there are convincing arguments against averaging across stimuli or subjects. Multivariate filters can be used to isolate an ERP component of interest without the averaging procedure. However, we would like to have certainty that the output of the filter accurately represents the component. *Methods:* We extended the linearly constrained minimum variance (LCMV) beamformer, which is traditionally used as a spatial filter for source localization, to be a flexible spatio-temporal filter for estimating the amplitude of ERP components in sensor space. In a comparison study on both simulated and real data, we demonstrated the strengths and weaknesses of the beamformer as well as a range of supervised learning approaches. *Results:* In the context of measuring the amplitude of a specific ERP component on a single trial basis, we found that the spatio-temporal LCMV beamformer is a filter that accurately captures the component of interest, even in the presence of both structured noise (e.g., other overlapping ERP components) and unstructured noise (e.g., ongoing brain activity and sensor noise). *Conclusion:* The spatio-temporal LCMV beamformer method provides an accurate and intuitive way to conduct analysis of a known ERP component, without averaging across trials or subjects. *Significance:* Eliminating averaging allows us to test more detailed hypotheses and apply more powerful statistical models. For example, it allows the usage of multi-level regression models that can incorporate between subject/stimulus variation as random effects, test multiple effects simultaneously and control confounding effects by partial regression.

3.2 Introduction

EEG measures the electrical activity that spreads outwards from its origin source through the various tissues and fluids in the head, until it is registered by the electrodes on the scalp [52]. This means that an EEG electrode typically picks up

a mixture of signals originating from many different sources in the brain. Likewise, signals originating from a single source are typically picked up by multiple electrodes. In many cases, researchers are interested in the behavior of a specific signal, which is but a single voice within the chatter of all the various processes going on in the brain. The focus in this paper is on isolating signal components that are part of the ERP [142]. These components are time-locked to perceptual, cognitive or motor events and correspond to specific peaks and valleys in the ERP waveform. They are named by their positive/negative deflection and time offset relative to the onset of the event in milliseconds (e.g., P300, N400). Such a component will be referred to here as an ERP COI. Of particular interest is the amplitude of a COI. In this study, we assume that increased activity at the neural generator responsible for the COI, translates into a uniform, linear scaling of the COI shape. Whenever we speak of the amplitude of the COI, we refer to this scaling.

3.2.1 Limitations of the averaging technique

A widely used technique to isolate the ERP is to extract trials (also referred to as epochs) from the ongoing EEG, based on the onsets of event markers, and compute their average. Through this operation, signal components that are time-locked to the event markers are retained while unrelated components are suppressed [54, 143]. Statistical analysis of COI amplitude typically proceeds by taking some measurement on the resulting waveform in a certain ROI, i.e.some specific electrodes and time range), usually the mean voltage [55]. We will refer to this method as the ROI-mean measure. A downside of this intuitive approach is that, due to the averaging procedure, it yields only a few data points per subject; usually one for each experimental condition.

Since so few data points are produced, studies that employ averaging across trials traditionally follow a design that manipulates a single property of the stimuli or task per experimental condition. Such an experimental design is limiting, as not only does it take time to test different manipulations one by one, but manipulating only a single property of a stimulus can be very difficult. For example, in semantic studies, constructing two word lists where the words differ in only one relevant property (e.g., length, frequency of use, age of acquisition, etcifnextchar...) is almost impossible [123].

By increasing the number of subjects, experimental designs become possible that enable the use of regression techniques when dealing with ERPs[77, 103, 144–147]. This opens up the possibility to test the effect of multiple manipulations simultaneously and allows correction for unwanted effects through partial regression. As hypotheses become more intricate and effect sizes become smaller, these designs require ever increasing amounts of subjects.

An additional disadvantage of averaging across stimuli or subjects is that statistical models are no longer able to incorporate either between-stimulus or between-subject variability. This is referred to as the language-as-a-fixed-effect fallacy and cannot be simply ignored [148, 149]. To address this fallacy, multilevel models, such as LME models are becoming increasingly popular in linguis-

tic studies [150, 151]. While these models can theoretically account for between-stimulus and between-subject effects[135], they must operate on unaveraged data to do so. Without averaging across trials, the mean or peak voltage in a ROI measure can only reveal very strong effects, given a large amount of subjects. For example, Vossen *et al.* [152] present a study where multi-level models are used for statistical analysis of the ERP on 85 subjects that were administered electrical pain stimuli.

3.2.2 Performance criteria for multivariate techniques

There is clearly an opportunity for methods that do not rely on averaging to isolate a COI. A filter that separates overlapping ERP components should boost signal-to-noise substantially. The application of such filters (e.g., [53, 109, 153]), falls under the category of techniques known as multivariate analysis. The usefulness of multivariate methods, especially linear ones, for EEG analysis has been acknowledged for a long time [154] and single-trial analysis has been growing in popularity [155].

Multivariate methods can be applied to produce filters that combine the EEG signal from multiple electrodes (i.e.a spatial filter), multiple time samples (referred to in this paper as a temporal filter) or both (a spatio-temporal filter) into one representative value. Since we are interested in estimating the amplitude of a COI in a trial, thus reducing the samples from all electrodes and all time points to a single value, we will be looking at spatio-temporal filters.

If the output of such a filter is used as estimation for the amplitude of a COI, two important performance criteria are that it:

1. correlates well with the actual amplitude of the COI (sensitivity)
2. does *not* correlate with any structured interfering signals, such as other ERP components or eye movements (specificity)

A filter that scores reasonably well on the sensitivity criterion does not necessarily score well on the specificity criterion. The specificity criterion states that it is preferable for any variation of the filter output that is not explained by the amplitude of the COI to be due to unstructured (e.g., zero-mean Gaussian) noise. Take for example a filter which output correlates with the amplitudes of both the N/P150 and N400 components. If a researcher uses this output as estimation for N400 amplitude, he will mistakingly find that the N400 is modulated both by word-frequency (that actually does modulate the N400 [7]) and font size (which modulates the N/P150, but not the N400 [58]).

3.2.3 The beamformer technique

In order to create a multivariate filter that both performs well on the criteria above and can be interpreted intuitively, we examined beamformer techniques. Beamformers were originally formulated for processing sonar, radar, and seismic data [156] and have since been applied to EEG as a spatial filter that isolates the signal originating from a specific point on the cortex [108]. They have

also been applied to BCIs to isolate activity from specific regions of the motor cortex [157], which allows a user to send commands to a computer system by imagining movement that activates these regions. In this paper, we bring a beamformer algorithm into the context of isolating a COI, even when its source location in the brain is unknown, by extending the original formulation to a spatio-temporal filter. A feature that makes beamformers compelling is that a template of the COI is given as explicit input to the algorithm. It then proceeds to construct a filter that isolates the COI by combining the template with the inversed covariance matrix of the entire signal. The method is very transparent, because the user is in full control of constructing the template of the COI and robust estimation of the covariance matrix for EEG signals is a thoroughly studied subject [158–160].

3.2.4 Assessment of various multivariate techniques

The appropriateness of a signal processing technique depends on the question the investigator would like to answer and the underlying assumptions placed on the data. In order to demonstrate the circumstances when a beamformer is suitable and when another multivariate method is preferable, a simulation study was performed. Simulated EEG recordings were generated with varying parameters, including the level of structured and unstructured noise, variation of the COI shape between subjects, etcifnextchar... The model was used to analyze the performance (in terms of the two criteria listed above) of the beamformer as well as the traditional ROI-mean measure and a variety of supervised learning approaches. The results demonstrate the strengths and weaknesses of each method when it comes to accurately isolating a COI from the ongoing EEG signal.

Finally, each method was applied to EEG data recorded during a semantic priming experiment [161]. Semantic priming is a commonly used method to study the operation and organization of the semantic processes in the brain (for a review, see [2, 7]). In such an experiment, the subject is given a task that involves responding to a target semantic stimulus, which in this study was a single word. When the target is preceded by a semantically and/or associatively related prime word, it allows the subject to respond more efficiently, lowering the RT of the subject in a decision task that requires reading the words [1, 25]. The semantic priming effect has also been successfully measured using electroencephalography (EEG), where it manifests itself in the ERP, mainly through a component called the N400 [7, 75]. A regression study was performed to analyze the relation between N400 amplitude and several properties of the stimuli. To avoid the language-as-a-fixed-effect fallacy, a linear mixed-effects (LME) model was employed, where both subjects and stimuli were entered as random effects, following the recommendations of Baayen *et al.* [135].

Table 3.1: Mathematical notation

m	number of EEG electrodes
n	number of time samples in a trial
r	number of trials in a recording
r_{train}	number of trials used for training
s	number of subjects that were recorded
y	the true COI amplitude during a trial
$\hat{y}$	unitless estimation of the COI amplitude
$\mathbf{S}$	$m \times n$ matrix containing the samples at all EEG electrodes and time points during a single trial
$\mathbf{A}$	$m \times n$ matrix containing the shape of the COI during a single trial
$\mathbf{N}$	$m \times n$ matrix containing the summation of all noise sources during a single trial
$\mathbf{x}$	$(mn) \times 1$ vector containing a "column-wise flattened" version of $\mathbf{S}$, constructed by stacking the columns of $\mathbf{S}$
$\mathbf{X}$	$(mn) \times r$ matrix composed by concatenation of r vectors $\mathbf{x}$
$\mathbf{w}$	$(mn) \times 1$ vector representing a spatio-temporal filter

3.3 Methods

3.3.1 Linear model of EEG

In this section, we introduce the mathematical model used to generate the artificial EEG data and to discuss the merits of the various multivariate methods. The mathematical notation adopts the convention of denoting variables that represent a scalar by cursive letters, those that represent vectors with bold lowercase letters and those that represent matrices by bold uppercase letters. See table 3.1 for a summary of all the variables and their meaning.

During a single trial, EEG signals are sampled at n time points at m electrodes, yielding an $m \times n$ matrix $\mathbf{S}$. An entire recording contains r trials, with corresponding matrices $\mathbf{S}_i,\ i = 1, \ldots, r$.

Activity in the brain is modeled as the summed activation of various source equivalent dipole generators in the brain (figure 3.1). The model assumes that activity at a generator is linearly transferred through volume conduction in the head to the EEG electrodes. When a generator is active, it produces an *activation pattern* on the sensors, as electrodes close to the generator will pick up more activity than electrodes further away (depicted as rectangles in figure 3.1, top), and also across time as the activity of the generator rises and decays (depicted as curves in figure 3.1, bottom). One of these generators is used to model the COI and will be referred to as the generator of interest (GOI). Other generators which activity is time-locked to the onset of the trial are regarded as structured noise. They are used to model ERP components that are not the COI. Generators which activity is not time-locked to the onset of the trial are regarded as unstructured noise. They are used to model task irrelevant EEG that would be suppressed

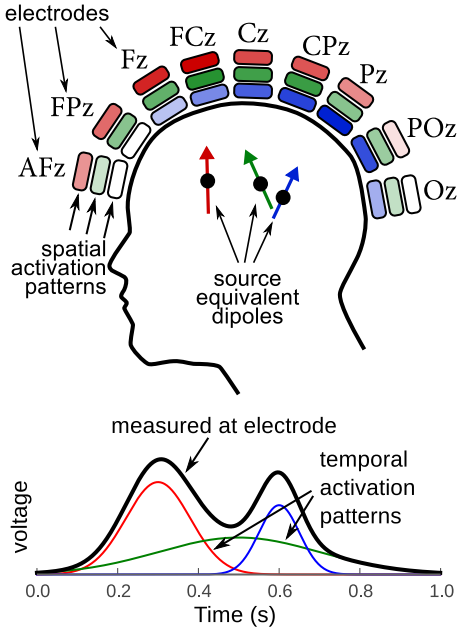


Figure 3.1: Sketch showing how signals generated at different dipole generators are captured by the EEG electrodes. Activity at the three generators is drawn in blue, red and green. Each electrode records a mixture of the three generators. (top) The spatial activation patterns produced by the generators are shown as rectangles with different color intensities. (bottom) The temporal activation patterns of the generators are shown as curves in different colors. The black curve represents the summation of these patterns as captured by a single EEG electrode.

by averaging over trials instead of showing as ERP components. The last noise source in the model is sensor noise, which is modeled as zero-mean Gaussian noise.

Let y denote the activity at the GOI during a trial. The activation pattern $\mathbf{A}$, an $m \times n$ matrix, maps activity at y to activity recorded both at the EEG electrodes (spatially) and to time samples (temporally). All other EEG activity during the trial (other ERP components, ongoing brain activity, sensor noise, etc.) is modeled as $m \times n$ matrix $\mathbf{N}$, which can be decomposed into structured noise $\mathbf{N}_{\text{struct}}$, unstructured noise $\mathbf{N}_{\text{unstruct}}$ and sensor noise ϵ:

$$\mathbf{S} = y\mathbf{A} + \mathbf{N}, \tag{3.1}$$

$$\mathbf{S} = y\mathbf{A} + \mathbf{N}_{\text{struct}} + \mathbf{N}_{\text{unstruct}} + \epsilon. \tag{3.2}$$

The activation pattern $\mathbf{A}$ can be interpreted as the shape of the COI and y can be interpreted as its amplitude. Each of the methods explored in this study takes a different approach to deduce the amplitude y of the COI, given the signal $\mathbf{S}$ recorded during a trial.

3.3.2 Multivariate filters

Linear multivariate filters aim to capture activity generated at the GOI by linearly combining EEG samples. These filters aim to eliminate activity generated by noise generators to isolate the GOI, which succeeds if the generators' dipoles differ in location and/or orientation and/or when the dipole's activity differs in timing. A spatial filter combines samples recorded at the same time, but at different electrodes, into one virtual component that optimizes a particular property of the signal, for example correlation of the output with activity of a single generator over time. In the example sketched in figure 3.1, a possible spatial filter would be the sample at Pz minus the sample at FPz. The Pz channel records the desired activity stronger than FPz and both electrodes capture the unwanted dipoles approximately equal, so the Pz−FPz combination would contain the desired signal with much less noise. A temporal filter combines samples recorded at the same electrode or virtual component, but at different time points, into one representative value that optimizes a particular property of the signal, for example correlation of the output with the amplitude of a COI. In the example, if we would be interested in obtaining an amplitude measurement of the blue component, the traditional approach would be to take the sample at 600 ms, or the mean of the samples from 500 ms to 700 ms. However, due to overlapping components, it would in this case be more accurate to take the sample at 600 ms, which measures both the blue and green components, and subtract the sample at 500 ms, which measures mostly the green component.

In this study, we are concerned with linear spatio-temporal filters that combine samples from both different electrodes and different time points. A linear spatio-temporal filter can be represented by a vector $\mathbf{w} \in \mathbb{R}^{(mn)\times 1}$ that operates on a data vector $\mathbf{x} \in \mathbb{R}^{(mn)\times 1}$ constructed by stacking all the columns of $\mathbf{S}$. The

result of such filtering is a scalar value $\hat{y}$:

$$\hat{y} = \mathbf{w}^{\mathsf{T}}\mathbf{x}, \tag{3.3}$$

which can be used as estimation for the amplitude of the COI during the trial (y). The optimal $\mathbf{w}$ depends on both the activation pattern of the COI $\mathbf{A}$ and the noise $\mathbf{N}$. Even if $\mathbf{A}$ is known in advance due to prior studies, it is difficult to know the various noise sources in advance. The most successful multivariate filters therefore contain data-driven elements to estimate $\mathbf{A}$ and/or $\mathbf{N}$ from the recording(s) currently under consideration.

3.3.3 Supervised learning approach

A popular way to estimate $\mathbf{w}$ employs supervised training. Linear regression models in particular have long been used for effective spatio-temporal filtering [111–113]. Since implementations of such algorithms are readily available, we skip the implementation details and describe how they can be applied to the problem at hand.

In the context of this study, the learning algorithms operate on a training set which consists of a set training pairs $\{(\mathbf{S}_i, y_i) : i = 1, \ldots, r_{\text{train}}\}$, where $\mathbf{S}_i$ contains the data for a single training trial, y_i is the true (or a good estimate of the) amplitude of the COI in trial $\mathbf{S}_i$, and r_{train} is the number of trials used for training. Given the training set, the learning algorithm will produce a weight vector $\mathbf{w}$, which is interpreted in this paper as a spatio-temporal filter, to estimate the amplitude $\hat{y}$ of the COI given trial data $\mathbf{S}$ in the form of vector $\mathbf{x}$ using (3.3).

In many cases, it is difficult to obtain reliable estimates of the exact amplitude of the COI to use as training labels y_i. A useful approximation in this case is to limit the training labels to $\{-1, 1\}$, encoding "small" and "large" COI amplitudes respectively. The training data is in this case limited to the trials expected to have an exceptionally small or large COI amplitude. During training, the original regression problem is now substituted by a classification problem, which aims to distinguish between two classes using a decision boundary.

Suitable training data can be obtained by creating a contrast study where the experimental manipulation is designed to only modify the amplitude of the COI. The quality of the training labels and therefore the quality of the filter depends on the suitability of the contrast used (see section 3.3.7 for our case).

In this study, we evaluated the linear support vector machine (lSVM) [162] as representative supervised learning algorithm. In addition to the straightforward approach of using all samples as features, three variations were evaluated. The first variation was to restrict the features to an ROI subset of the data (restricted-lSVM), where the ROI was the same as the one used for the ROI-mean measurement. The second variation was to train the lSVM on the training data, pooled from all recordings (group-lSVM). The final variation was to first pass the data through a spatial filter, using the xDAWN algorithm [116], keeping the 4 most descriptive spatial components, and then using an lSVM to perform the final filtering. The covariance matrix for the xDAWN algorithm was estimated using oracle approximating shrinkage (OAS) [158]. In all cases, the penalty parameter (C) for

the lSVM was optimized using five-fold cross-validation on the training set. A further assortment of alternative supervised learning approaches are evaluated in the supplementary information section.

3.3.4 Beamformer approach

In contrast to the supervised learning approach, beamformers take the shape of the COI (matrix $\mathbf{A}$) as explicit input. Of the various beamformer approaches, the LCMV beamformer [108] seems suitable for our purposes as it was shown to accurately recover the activity at the GOI over time [122].

The LCMV beamformer was originally formulated as a spatial filter $\mathbf{w}_{\text{sp}} \in \mathbb{R}^{m\times 1}$ that, when applied to the centered EEG signal $\mathbf{S}$, minimizes the variance of the result $\mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S}$:

$$\mathbf{w}_{\text{sp}} = \arg\min_{\mathbf{w}_{\text{sp}}} \mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S}(\mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S})^{\mathsf{T}} = \arg\min_{\mathbf{w}_{\text{sp}}} \mathbf{w}_{\text{sp}}^{\mathsf{T}}\Sigma_{\text{sp}}\mathbf{w}_{\text{sp}}, \tag{3.4}$$

where $\Sigma_{\text{sp}} \in \mathbb{R}^{m\times m}$ is the spatial covariance matrix of the signal $\mathbf{S}$, which can be replaced by $\sum_{i=1}^{r} \text{cov}\mathbf{S}_i$. To estimate this covariance matrix, one might consider using robust estimation techniques that employ shrinkage [158–160]. In this study, as said above, all covariance matrices were estimated using OAS [158].

To avoid trivial solutions of (3.4), vector $\mathbf{w}_{\text{sp}}$ must be constrained appropriately, for example, linearly:

$$\mathbf{a}_{\text{sp}}^{\mathsf{T}}\mathbf{w}_{\text{sp}} = 1, \tag{3.5}$$

where $\mathbf{a}_{\text{sp}} \in \mathbb{R}^{m\times 1}$ is the spatial activation pattern of the GOI (depicted by the colored rectangles in figure 3.1, top).

Following [108], the solution of (3.4) can be found using for example the method of Lagrange multipliers:

$$\mathbf{w}_{\text{sp}} = \frac{\Sigma_{\text{sp}}^{-1}\mathbf{a}_{\text{sp}}}{\mathbf{a}_{\text{sp}}^{\mathsf{T}}\Sigma_{\text{sp}}^{-1}\mathbf{a}_{\text{sp}}}. \tag{3.6}$$

The formulation of the LCMV beamformer can be expanded to a spatio-temporal filter. Let $\mathbf{X} \in \mathbb{R}^{(mn)\times r}$ be a matrix consisting of r columns $\mathbf{x}_i$, which are the column-wise flattened versions of the corresponding EEG trials $\mathbf{S}_i$ $(i = 1,\ldots,r)$, $\Sigma \in \mathbb{R}^{(mn)\times(mn)}$ be the covariance matrix of $\mathbf{X}$, and $\mathbf{a} \in \mathbb{R}^{(mn)\times 1}$ be a vector containing the column-wise flattened version of $\mathbf{A}$. Similarly to the spatial case, the spatio-temporal linearly constrained minimum variance (stLCMV) $\mathbf{w} \in \mathbb{R}^{(mn)\times 1}$ is the result of minimization of the variance of $\mathbf{w}^{\mathsf{T}}\mathbf{X}$ constrained by $\mathbf{a}^{\mathsf{T}}\mathbf{w} = 1$:

$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{a}}{\mathbf{a}^{\mathsf{T}}\Sigma^{-1}\mathbf{a}}. \tag{3.7}$$

Alternatively, a simpler spatio-temporal filter can be obtained by sequentially applying to the trial data $\mathbf{S}$ a spatial beamformer $\mathbf{w}_{\text{sp}}$ and then a temporal

beamformer $\mathbf{w}_{\text{tmp}}$. To define $\mathbf{w}_{\text{tmp}}$, we define $\mathbf{B} \in \mathbb{R}^{r \times n}$ as a matrix containing the results of applying the spatial beamformer $\mathbf{w}_{\text{sp}}$ to the EEG trials $\mathbf{S}_i$:

$$\mathbf{B} = \begin{pmatrix} \mathbf{w}_{\text{sp}}^{\top} \mathbf{S}_1 \\ \mathbf{w}_{\text{sp}}^{\top} \mathbf{S}_2 \\ \vdots \\ \mathbf{w}_{\text{sp}}^{\top} \mathbf{S}_r \end{pmatrix}, \tag{3.8}$$

the covariance matrix of which we denote by Σ_{tmp}. Considering $\mathbf{a}_{\text{tmp}} \in \mathbb{R}^{n \times 1}$ as a vector containing the temporal activation pattern of the GOI (e.g., one of the curves in figure 3.1, bottom), the temporal LCMV beamformer can be expressed as:

$$\mathbf{w}_{\text{tmp}} = \frac{\Sigma_{\text{tmp}}^{-1} \mathbf{a}_{\text{tmp}}}{\mathbf{a}_{\text{tmp}}^{\top} \Sigma_{\text{tmp}}^{-1} \mathbf{a}_{\text{tmp}}}. \tag{3.9}$$

The spatial and temporal beamformers may be chained together to perform spatio-temporal filtering:

$$\hat{y} = \mathbf{w}_{\text{sp}}^{\top} \mathbf{S} \mathbf{w}_{\text{tmp}}, \tag{3.10}$$

to obtain a single scalar value $\hat{y}$ from an EEG trial $\mathbf{S}$.

The resulting filter looses the advantage of being able to take interactions between electrodes over time (e.g., a traveling wave) into account, but the number of free parameters is greatly reduced. Where the original spatio-temporal filter had mn (in this study $32 \times 54 = 1728$) free parameters, chaining a separate spatial and temporal filter has $m + n$ (in this study $32 + 54 = 86$). This beamformer will be referred to as the chained LCMV beamformer (chained-LCMV).

3.3.5 Modeling the activation pattern

While the covariance matrix can be readily computed from the data, it is up to the researcher to provide the beamformer filter with the activation pattern of the GOI: the shape of the COI. In the traditional application of beamformers, the problem of source localization, the activation pattern is computed through a realistic anatomical model of the subject's brain and head [108]. However, since uncovering the location of the GOI in the brain is not our purpose in this study, an anatomical model is not required.

We propose to estimate the activation pattern through the traditional manner of averaging across trials, using data from all available recordings. First, a training set is selected analogous to the one used in section 3.3.3. The training pairs in the training set were divided into two classes: those known to contain the COI with a large amplitude and those with a small COI amplitude (see section 3.3.7 for the contrast we used for the evaluation on real EEG data). Grand average ERPs are constructed for both classes and the difference wave is taken as first estimate of the activation pattern. Let matrix $\hat{\mathbf{A}} \in \mathbb{R}^{m \times n}$ denote this activation pattern, which can be interpreted as a template for the COI.

This template can be refined at will by the researcher. In this study, we opted to do this refinement by approximating the activation pattern $\hat{\mathbf{A}}$ as a product of the separate spatial $\hat{\mathbf{a}}_{\text{sp}}$ and temporal $\hat{\mathbf{a}}_{\text{tmp}}$ activation patterns. For the spatial pattern, we used the column of $\hat{\mathbf{A}}$ corresponding to the time point at which the COI reaches its maximum amplitude. To find it, a suitable ROI was first defined by visual inspection of $\hat{\mathbf{A}}$. Let $\mathbf{c} \subset \{1,\ldots,m\}$ denote a set of row indices of $\hat{\mathbf{A}}$ corresponding to the electrodes of interest. Likewise, let $\mathbf{t} \subset \{1,\ldots,n\}$ denote a set of column indices of $\hat{\mathbf{A}}$ corresponding to the samples from the time window of interest. Then:

$$t_{\text{peak}} = \arg\max_{t \in \mathbf{t}} \left| \sum_{c \in \mathbf{c}} \hat{\mathbf{A}}(c,t) \right|, \tag{3.11}$$

$$\hat{\mathbf{a}}_{\text{sp}} = [\hat{\mathbf{A}}(1,t_{\text{peak}}),\ldots,\hat{\mathbf{A}}(m,t_{\text{peak}})]^{\mathsf{T}} \in \mathbb{R}^{m\times 1}, \tag{3.12}$$

To construct the temporal activation pattern $\hat{\mathbf{a}}_{\text{tmp}} \in \mathbb{R}^{n\times 1}$, for each recording, a spatial beamformer $\mathbf{w}_{\text{sp}}$ was constructed from $\hat{\mathbf{a}}_{\text{sp}}$ using (3.6) and applied to the data:

$$\hat{\mathbf{a}}_{\text{tmp}} = \frac{1}{r} \sum_{i=1}^{r} \mathbf{S}_i^{\mathsf{T}} \mathbf{w}_{\text{sp}}. \tag{3.13}$$

The resulting temporal activation patterns $\hat{\mathbf{a}}_{\text{tmp}}$ were averaged across recordings. Finally, all values outside the range defined by the first and last zero-crossings inside the temporal ROI (from 300 ms to 600 ms) were set to zero to eliminate small deviations from zero at irrelevant time points. The refined spatio-temporal template for the COI is then a product of the spatial and temporal activation patterns:

$$\hat{\mathbf{A}}_{\text{refined}} = \hat{\mathbf{a}}_{\text{sp}} \hat{\mathbf{a}}_{\text{tmp}}^{\mathsf{T}}. \tag{3.14}$$

3.3.6 Evaluation on simulated data

In order to evaluate the different methods of measuring COI amplitude, simulated EEG recordings were generated. The software model allows control over the different signal components listed in (3.2): y, $\mathbf{A}$, $\mathbf{N}_{\text{struct}}$, $\mathbf{N}_{\text{unstruct}}$ and ϵ.

For a pool of 10 subjects, 400 trials were simulated that consisted of virtual recordings at 32 electrodes, distributed over the scalp using the extended 10–20 system. Each simulated trial lasted 1 s and was sampled at 50 Hz, mimicking the properties of the real EEG data discussed later on.

The data were generated by simulating dipoles in a spherical head model. The activity at the dipoles was modeled as Gaussian curves:

$$f(t) = y\, e^{-\frac{1}{2}\left(\frac{t-t_{\text{peak}}}{\sigma}\right)^2}, \tag{3.15}$$

where $f(t)$ is the activity (in μV) at the dipole at time t, y is the amplitude of the dipole, the peak activity occurs at t_{peak} and σ determines the speed at which the activity reaches its peak and decays back to zero. For each time point, the spatial

activation pattern of each dipole was computed using a three layer bounded element method (BEM) model, where the layers represented the inner skull, outer skull and outer skin boundaries [163]. Each dipole has a location $\mathbf{l} = [l_x, l_y, l_z]$ defined as a x-, y- and z-coordinate, restricted to be within the inner skull layer of the BEM model, and an orientation $\mathbf{o} = [r_x, r_y, r_z]$ defined by Euler rotations around the x-, y- and z-axes.

The COI was simulated as a single dipole (the GOI) at a fixed location $\mathbf{l}_{\text{GOI}} = [0, 0, 0.5]$. Coordinates are normalized so the inner skull sphere has a radius of 1. This places this dipole centered between the auricular points (x-axis), centered between the nasion and inion (y-axis) and raised towards the top of the head (z-axis). Its orientation was $\mathbf{o}_{\text{GOI}} = [\pi, 0, 0]$ (pointing straight up), yielding a spatial activation pattern that centers on the Cz electrode, spreading radially to all other electrodes. The COI peaked at $t_{\text{peak}} = 0.4 + j$, where j is randomly drawn from a uniform distribution defined over the range $[-J_{\text{COI}}, J_{\text{COI}}]$ for each recording, but held constant between trials. The parameter J_{COI} is the first model parameter: the amount of temporal jitter of the COI between subjects. The width of the COI was held constant at $\sigma = 0.05$. For half of the trials, the amplitude y of the COI was 1 µV and for the remaining trials it was 0 µV.

The sensor noise ϵ was modeled for each sample individually, by randomly drawing from a zero-mean, Gaussian distribution with a standard deviation of 0.1.

The structured noise $\mathbf{N}_{\text{struct}}$ was modeled using 20 dipoles with initially random parameters: each parameter value was drawn from a uniform distribution, defined over the range of possible valid values for the parameter. Between trials, all parameters were held constant except for the amplitude of the dipole activity. For each trial, the amplitude parameter y was randomly drawn from a uniform distribution over the range $[0, S_{\text{struct}}]$, where S_{struct} is the second model parameter: the scale of the structured noise.

Finally, the unstructured noise $\mathbf{N}_{\text{unstruct}}$ was modeled using 20 dipoles in the same manner as the structured noise dipoles, except that parameters for the unstructured noise dipoles were randomly drawn for each individual trial: no parameter was held constant. The uniform distribution used for the amplitude parameter was defined over the range $[0, S_{\text{unstruct}}]$, where S_{unstruct} is the third model parameter: the scale of the unstructured noise.

The task for each of the filters was to estimate the amplitude of the COI (y) for each trial, which corresponds to the peak amplitude of the GOI dipole in the model and to y in (3.2). The total set of trials was split 50-50 into a training and test set. Each set contained 200 trials for each of the 10 subjects, 100 trials containing the COI with an amplitude of 1 µV and 100 trials not containing the COI (its amplitude was 0 µV). Training labels l were produced using a mixture of the true amplitude y of the COI and the amplitude y_{struct} of one of the structured noise dipoles:

$$l = R_{\text{labels}}\, y + (1 - R_{\text{labels}})\, y_{\text{struct}}, \tag{3.16}$$

where R_{labels} is the fourth model parameter: the accuracy of the training labels, ranging from 0 (labels follow a noise component) to 1 (labels follow the COI).

Each of the multivariate filters was trained on the training set, plus the corresponding training labels **l**. Each method then produced an estimate of COI amplitude ($\hat{y}$) for each of the trials in the test set, for which the training labels were withheld.

The entire procedure was ran 10 times in order to assess the variation between simulation runs, producing 100 data sets (10 subjects × 10 runs). The performance of each filter was observed for different values of the four model parameters J_{COI}, S_{struct}, S_{unstruct} and R_{train}. The base model settings were $J_{\text{COI}} = 0$, $S_{\text{unstruct}} = 1$, $S_{\text{struct}} = 3$, $R_{\text{labels}} = 1$. During the simulation, each parameter was changed in isolation, leaving the others at their base values.

3.3.7 Evaluation on real EEG data

The analysis on real EEG data was conducted on the dataset recorded in [161], where the COI is the N400 potential. Subjects read a series of sequentially presented words, organized in pairs, and pressed one of two buttons to indicate whether the two words of a word-pair were related or not. The prime word was presented for 200 ms and the target-word for 1000 ms with a stimulus onset asynchrony (SOA) of 500 ms. Since a speeded button response task will generate ERP components that can mask N400 modulations [161], the subjects performed delayed their button response until a cue was given.

The experiment was performed with 10 university students (3 female, aged 19–27 years), all right-handed and native speakers of Flemish-Dutch. Ethical approval of this study has been granted by an independent ethical committee ("Commissie voor Medische Ethiek" of the UZ Leuven). This study was conducted according to the most recent version of the declaration of Helsinki.

The word-pairs used were a total of 800 Flemish-Dutch word-pairs, selected with varying FAS, as determined from an association norm dataset [46]. In this norm dataset, FAS between a prime and target word is defined as the number of respondents, out of 100, that wrote down the target as first response to the prime word in a free association task. The list of word-pairs consisted of the top 100 strongest related word-pairs (FAS ranged 69–95, mean FAS = 75.62) and 100 word-pairs where the prime and target words were randomly chosen and no record of the word-pair existed in the association norm data, therefore having an assumed FAS of zero. The remaining 600 word-pairs were chosen such that the logarithm of their FAS score is uniformly distributed using a log scale. The log scale was chosen because when the association norm data were analyzed, some properties of the word-pairs that co-vary with the FAS, correlate better with its logarithm than the raw values [161]. All selected words for the stimulus list have a length of 4–6 letters, a minimum word frequency of 2 occurrences per 10^6 words in the SUBTLEX-NL corpus [164] and a minimum in-degree of 5 in the association norm dataset.

In addition to capturing the button response of the participant, EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System, having a 5th order frequency filter with a pass band from 0.16 Hz to 100 Hz, and sampled at 2048 Hz. An EOG was recorded simul-

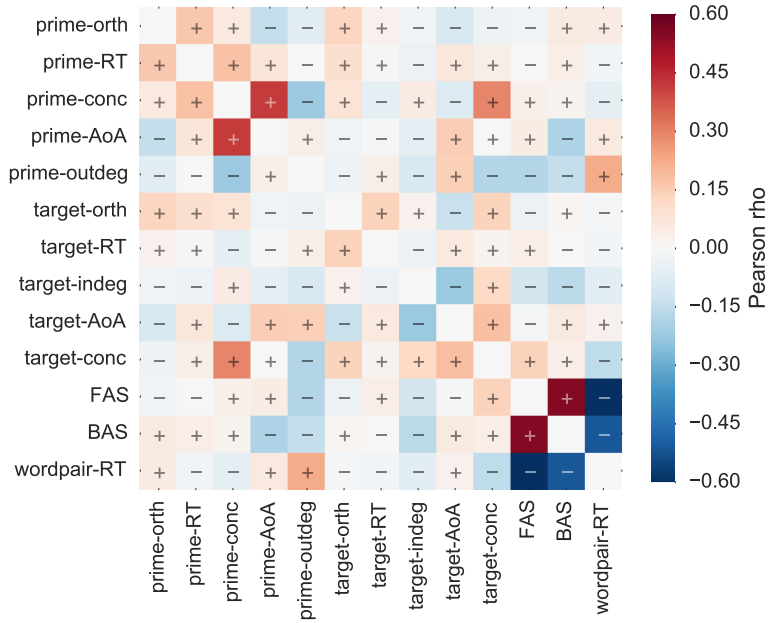


Figure 3.2: Pearson's correlation coefficients between dependent variables used in this study. See table 3.2 for a description of each one. The sign of coefficients is indicated with '+' and '-' symbols.

taneously and used to reduce eye artifacts in the EEG using the procedure outlined in [131]. Two electrodes were placed on both mastoids and their average was used as a reference for the EEG.

Stimulus properties

Since the true amplitude of the N400 is unknown, performance was based on a regression analysis with a selection of stimulus properties that have been shown to correlate with semantic priming in earlier regression studies [77, 85, 147]. See table 3.2 for a complete list and explanation of the stimulus properties used in the regression analysis. The aggregated set covers some strong and weaker correlates with the N400. Pearson's correlation coefficients between all stimulus properties are presented in figure 3.2.

Data preprocessing

The EEG was bandpass filtered offline between 0.1 Hz and 50 Hz by a 4th order zero-phase Butterworth filter to attenuate large drifts and irrelevant high frequency noise, but retain eye movement artifacts. The EOG was used to attenuate eye artifacts from the EEG signal using the regression method outlined in [131]. After the EOG correction procedure, the signal was band pass filtered again using a tight passband around the frequency range in which the N400

Table 3.2: Stimulus properties used in the regression study

Property	Description	Ref.
prime-orth	The orthographic neighborhood size of the prime word, i.e.the number of valid Dutch words with a Levenshtein distance of 1 from the prime word. Calculated using the SUBTLEX-NL corpus.	[164]
prime-RT	Reaction time to the prime word in a lexical decision task.	[81]
prime-outdeg	Number of outgoing links of the prime word, with an association strenght of ≥ 2 in the association norm data.	[46]
prime-AOA	Age of acquisition rating for the prime word.	[165]
prime-conc	Concreteness rating for the prime word.	[165]
target-orth	The orthographic neighborhood size of the target word, i.e.the number of valid Dutch words with a Levenshtein distance of 1 from the target word. Calculated using the SUBTLEX-NL corpus.	[164]
target-RT	Reaction time to the target word in a lexical decision task.	[81]
target-outdeg	Number of outgoing links of the target word, with an association strenght of ≥ 2 in the association norm data.	[46]
target-AOA	Age of acquisition rating for the target word.	[165]
target-conc	Concreteness rating for the target word.	[165]
FAS	The logarithm of the forward association strength between the prime and target words.	[46]
BAS	The logarithm of the backward association strength between the prime and target words.	[46]
wordpair-RT	The mean response time of the subjects to the word-pair in a speeded button response task, obtained during a separate recording session several months prior to the current experiment.	[161]

component was found, namely between 0.5 Hz and 15 Hz, by a 4th order zero-phase Butterworth filter. Individual trials were obtained by cutting the continuous signal from 0.1 s before the onset of each target stimulus to 1.0 s after. Baseline correction was performed using the average voltage in the interval before the stimulus onset (−100 ms to 0 ms) as baseline value. Before applying any multivariate analysis methods, the signal was further downsampled to 50 Hz to reduce the dimensionality.

Training data

To construct the training set for the supervised learning algorithms and model the activation pattern for the beamformer, a contrast has to be created that will produce both trials with a low and trials with a high N400 amplitude. For the training set, the 100 word-pairs with an FAS of 0 were chosen as the low N400 amplitude condition, and 100 word-pairs with the highest FAS for the high N400 amplitude condition. This contrast in FAS is well known to produce clear differences in N400 amplitude [7, 12, 83]. The trained filters were then applied to the remaining trials.

Discarding the training set would mean a limited range of FAS for the trials in the test set, which would potentially eliminate a large portion of the N400 effect. Therefore, COI amplitudes were estimated for the training set as well, by using leave-one-out cross-validation scheme.

Statistics

Regression analysis between stimulus properties and N400 amplitude was done by means of a linear mixed-effects (LME) model. Since the stimulus properties used as independent variables are intercorrelated (figure 3.2) we have chosen for a univariate approach. Each independent variable is regressed onto the dependent variable in a separate model. All variables were z-transformed, so regression weights represent estimates of the Pearson correlation between the independent and dependent variable. Each regression model was fitted two times, one with both subjects (modeling slopes only) and word-pairs (modeling intercepts only) as random effects, and one with only word-pairs as random effects. If the first model did not achieve a significantly better fit than the latter, as measured using the area under curve (AUC) metric, the latter model was used. Models were fitted using maximum likelihood (ML) for computing the AUC metric, and using restricted maximum likelihood (REML) for computing t-values. This design follows the recommendations of Baayen *et al.* [135]. Degrees of freedom are hard to compute for mixed models and are often in the order of several thousands. Satterthwaite's method [136] was used to estimate them. Degrees of freedom are not provided in this text, as the relationship between t-scores and p-values converges at this number of degrees of freedom.

3.3.8 Software

A full description of the various software packages used in this study and an implementation of the beamformer methods can be found in the supplementary information.

3.4 Results

The performance of each multivariate filter on the datasets was assessed, as well as the traditional ROI-mean method.

Different ROI's were tried, and the best performing time range and electrode selections for the ROI-mean and restricted-lSVM methods was selected, namely for the simulation study electrodes FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s) and for the real EEG data electrodes CP5, CP6, P3, Pz, P4, PO3 and P04, time range 0.3 s to 0.5 s.

3.4.1 Simulated data

By modifying the four model parameters, artificial EEG datasets were generated with different properties. The performance of each method was assessed based on the two criteria listed in section 3.2.2 (figure 3.3). The first measure was the amount of variance of its output explained by the true amplitude of the COI, measured by regressing the estimated amplitude onto the true amplitude and computing the R^2 metric. Adhering to the sensitivity performance criterion, higher is better for this metric. The second measure was the amount of variance in the filter's output that could be explained by activity of the structured noise dipoles. This was quantified by multivariate-regression of the activity of the 20 noise dipoles onto the method's estimation of the COI amplitude and computing the R^2 metric. This measures how well the method manages to reduce the influence of nearby noise sources on its COI amplitude estimate. Adhering to the specificity criterion, lower is better for this metric. In the simulation, any variance in the method's output not explained by the true amplitude of the COI or structured noise must necessarily be due to unstructured noise.

The traditional ROI-mean is generally the worst performing method, as the method does not actively counteract noise sources. Without averaging, the ROI-mean method does not measure solely the amplitude of the COI, but a mixture of the COI and the surrounding structured noise sources.

As the unstructured noise amplitude (S_{unstruct}) increases, all methods start

Figure 3.3 ***(facing page)*****: Performance of various techniques to estimate COI amplitude under various simulated conditions.** The performance is quantified by two metrics: the variance of the output explained (R^2 stat) by the true amplitude of the COI (top, higher is better) and the variance explained by least squares regression with the structured noise sources (bottom, lower is better). Each curve represents an average of the analysis of 100 data sets (10 subjects × 10 runs). For clarity, confidence intervals of the curves are omitted in this figure. They can be found in the supplementary information.

ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
xDAWN+lSVM
Variance explained by true COI amplitude
Variance explained by structured noise
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10
S_{unstruct}
simulating interfering ongoing EEG
0 5 10 15 20 25 30 35 40
S_{struct}
simulating interfering ERP components
0.00 0.05 0.10 0.15 0.20
J_{signal}
temporal jitter of COI peak (s)
0.0 0.2 0.4 0.6 0.8 1.0
R_{labels}
accuracy of training labels

failing to properly isolate the COI (figure 3.3, first column). This is to be expected, as this type of noise is spherical, so there is no possible orthogonal linear projection. The lSVM (with and without the xDAWN spatial filtering step) fails somewhat faster than the other multivariate methods, because the quality of the training data diminishes and it starts overfitting on the unstructured noise. Limiting the amount of features (restricted-lSVM) or adding more training data from other recordings (group-lSVM) effectively counters this behavior. The beamformer methods perform on par with the restricted-lSVM and the group-lSVM. The ROI-mean method initially correlates highly with the unstructured noise, because it doesn't actively counteract it. As the unstructured noise increases, this high correlation disappears as the output now correlates more with the unstructured noise.

A filter's ability to isolate the COI from structured noise sources is dependent on its ability to successfully model both of them (figure 3.3, second column). The lSVM, restricted-lSVM and stLCMV methods are remarkably capable of countering structured noise sources, even when the stimulated structured noise (S_{struct}) is raised to 40 times the amplitude of the COI. The estimation technique for the template, used by the beamformer approaches, suffered in these extreme conditions, causing the performance of the stLCMV beamformer to drop slightly. As the structured noise increases, the group-lSVM starts failing, because it cannot adapt to the different structured noise dipoles in each recording. Instead, it has to find a solution that isolates the COI from all structured noise dipoles on all 10 recordings that where pooled together. The chained-LCMV is able to adapt to each individual recording, but still fails, probably due its trade-off between model simplicity and power. This is also the case for the xDAWN spatial filter.

As the structured noise increases, its influence on the filters' output is more severe, causing most of the methods to fail the specificity criterion: fluctuations in one of the other ERP components is influencing the estimation of COI amplitude. The stLCMV beamformer is the notable exception, as the correlation between its output and the noise components is unaffected by their amplitude.

For the methods that pool together multiple recordings (i.e. chained-LCMV, stLCMV and group-lSVM), differences in the shape of the COI across recordings is problematic (figure 3.3, third column). This is especially true for the beamformer methods, as the method we used for constructing the template did not take between-subject variability into account. The beamformers are designed to fail when the given template of the COI doesn't match its actual shape.

Since the output of supervised learning algorithms is closely tied to the training labels, it is sensitive to their accuracy (figure 3.3, last column). When other noise components correlate with the training labels in addition to the COI, the lSVM is unable to separate them, resulting in a decreasing correlation with the actual COI and an increasing correlation with the noise components. This can be alleviated by restricting the features to exclude some noise components (restricted-lSVM) or by pooling together more data, which increased the overall reliability of the training labels in this case (grouped-lSVM). The method used in this study to design the template for the COI used by the beamformers, also relies on the training labels. However, the beamformer methods are quite robust

to inaccuracies in the training labels, as data are pooled across recordings and values outside the ROI were explicitly set to zero in the template. Of particular note is the fact that even when the training labels follow one of the structured noise sources instead of the COI, the correlation between beamformers' output and the structured noise sources remains low. Finally, since the ROI-mean method does not rely on training labels, its output is unaffected.

3.4.2 Real EEG data

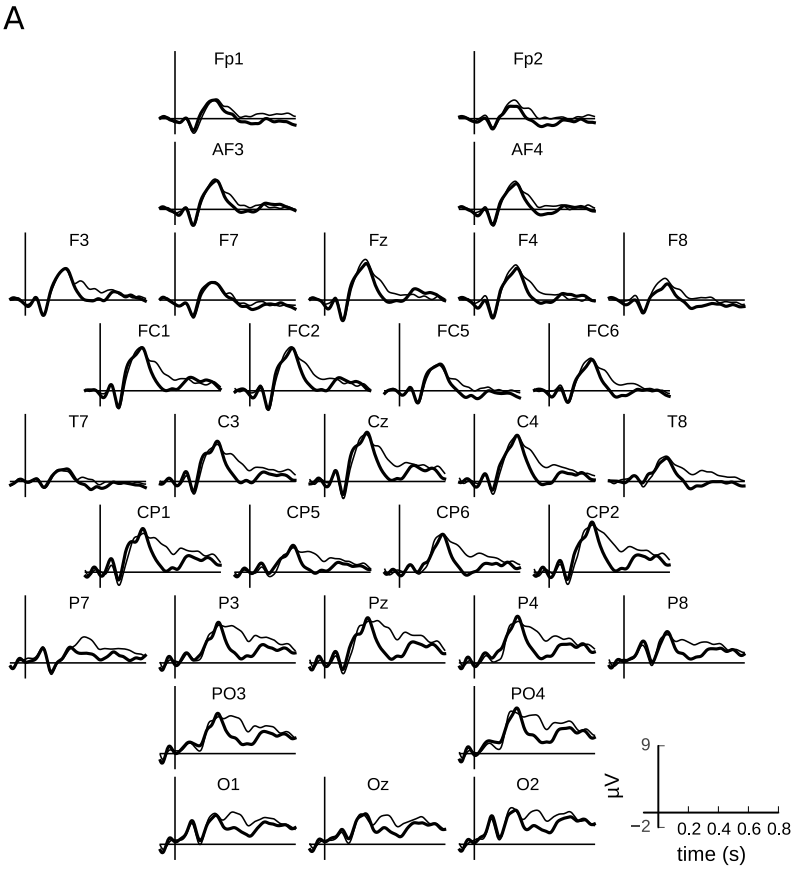
The grand average ERPs (figure 3.4) show clear N2 and P2 components, associated with the response to a visual stimulus. The N400 components is modulated by the FAS of the word-pairs as expected, growing in amplitude as the FAS decreases, peaking around 400 ms and strongest at the central-parietal electrodes, which is consistent with the literature [7].

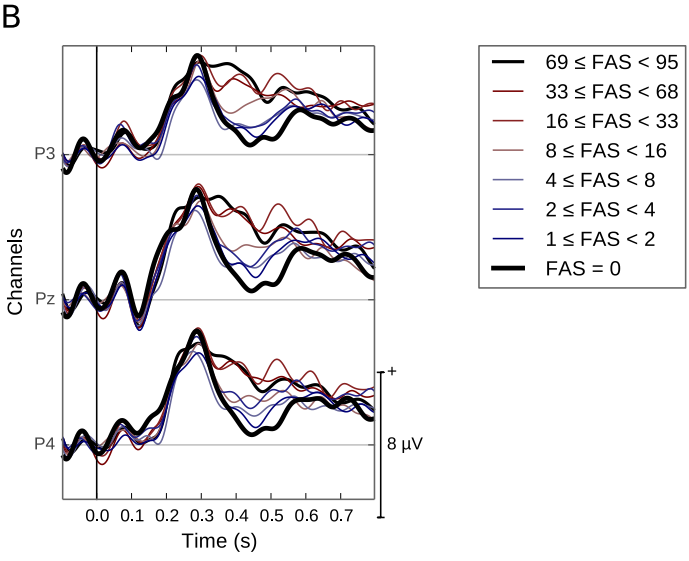
To test the performance of the multivariate filters in a more realistic setting, they were compared on a real EEG dataset, recorded in a semantic priming study [161]. A univariate regression study was performed using the amplitude of the N400, as quantified by the various analysis methods, as dependent variable and several stimulus properties (table 3.2), known to correlate with N400 amplitude, as independent variables (figure 3.5). To asses the performance of each method, we regarded the number of effects that were successfully found and the relative size of the regression weights. As all variables were z-transformed before entering them into the model, the regression weights can be interpreted like one would a Pearson's correlation coefficient. A comprehensive table of the exact regression weights, confidence intervals and p-values can be found in the supplementary information. Note that we have chosen not to correct the p-values for family-wise error rate. The purpose of this regression study is to compare the different analysis methods and due to the large number of tests, no significant effects would remain. Each stimulus property has already been shown in independent studies to correlate with the N400, so the occurrence of false-positives is unlikely. Also note that the various stimulus properties are correlated with each-other (figure 3.2), so the regression weights do not necessarily indicate the unique contribution of each stimulus property.

The spatial and temporal activation patterns used by the beamformer methods (created using the procedure described in section 3.3.5) are shown in figure 3.6.

The mean-ROI method only managed to find the two strongest effects on the N400: FAS and RT to the word-pair, but failed to find effects that strongly

Figure 3.4 ***(facing page)*****: Grand average ERPs in response to word-pairs with increasing FAS during the delayed condition.** Word-pairs were sorted by their FAS and grouped into 8 bins of 100 pairs and the average response is shown aligned to the onset of the target word (black vertical line). (a) ERPs of the first and last bins for each channel. This shows the most extreme case of unrelated word-pairs versus strongly related ones. Intermediate bins have been omitted here for clarity. (b) Blowup of three electrodes showing the ERP of each bin.


A
Fp1
Fp2
AF3
AF4
F3
F7
Fz
F4
F8
FC1
FC2
FC5
FC6
T7
C3
Cz
C4
T8
CP1
CP5
CP6
CP2
P7
P3
Pz
P4
P8
PO3
PO4
O1
Oz
O2
9
µV
−2
0.2 0.4 0.6 0.8
time (s)


B
69 ≤ FAS < 95
33 ≤ FAS < 68
16 ≤ FAS < 33
8 ≤ FAS < 16
4 ≤ FAS < 8
2 ≤ FAS < 4
1 ≤ FAS < 2
FAS = 0
Channels
P3
Pz
P4
+
8 µV
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Time (s)

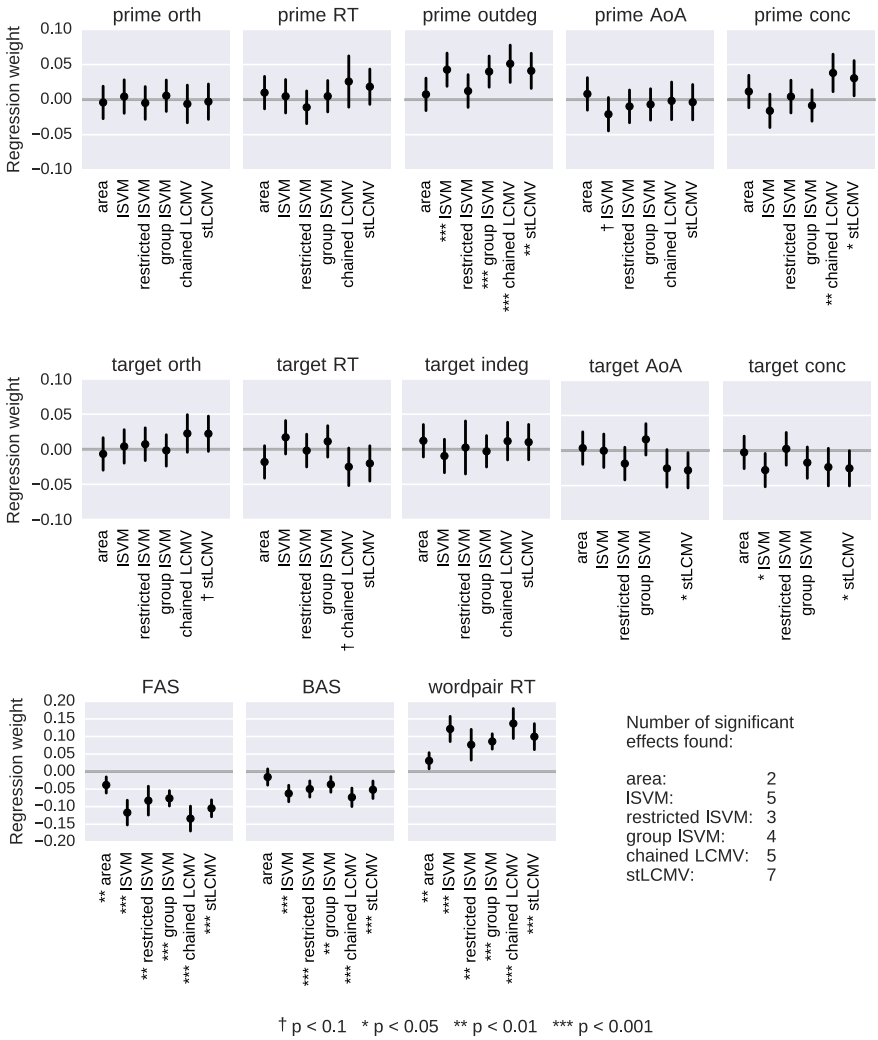


Figure 3.5: Univariate regression between each analysis method and each stimulus property. Dots indicate the regression weights obtained through the LME model and vertical lines indicate the 95% confidence interval. *P*-values are given for the null-hypothesis that the regression coefficient is zero. All variables were *z*-transformed before being entered into the model, hence regression weights can be interpreted as Pearson's correlation coefficients.

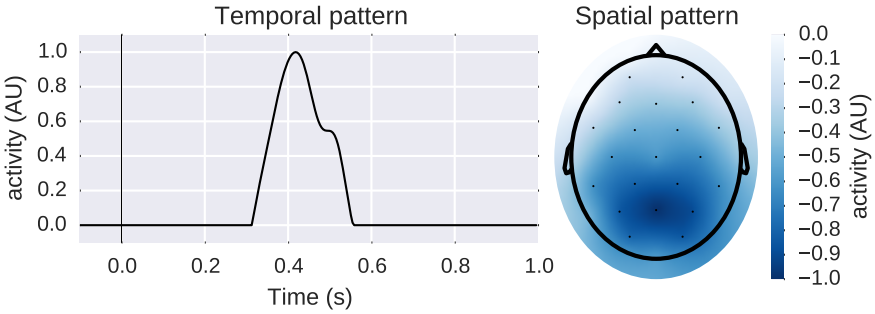


Figure 3.6: Temporal (left) and spatial (right) activation patterns used as template for the COI during beamformer analysis. In the visualization of the spatial pattern, dots represent the position of the electrodes and splines were used to interpolate the values between electrodes.

covariate with them (figure 3.2), namely BAS and the out-degree of the prime stimulus. Of the multivariate methods, only the restricted-lSVM failed to identify the latter effect. Where the chained-LCMV beamformer managed to find the effect of prime concreteness, the lSVM (with and without the xDAWN spatial filter) finds the effect of target concreteness instead. Finally, the stLCMV beamformer identified all of the effects the other methods found, as well as the effect of age-of-acquisition of the target word.

In terms of the size of the regression weights, the chained-LCMV beamformer performs best, followed by the lSVM, stLCMV beamformer, xDAWN+lSVM, restricted-lSVM, group-lSVM, and finally ROI-mean.

3.5 Discussion

The traditional method of measuring COI amplitude, namely taking the mean over a suitable ROI, is extremely sensitive to fluctuations of nearby (and not so nearby) ERP components (figure 3.3, bottom row). Averaging over many trials or subjects reduces these fluctuations and can vastly improve the result. However, when no such averaging is performed, it would be false to claim that limiting the analysis to a few selected electrodes and time points provides a measurement of the amplitude of the COI and the COI alone. Due to volume conduction, any electrode picks up signals from almost any part of the brain (figure 3.1, top) and the temporal dynamics of ERP components usually overlap as well (figure 3.1, bottom). The presence of other ERP components introduces a large amount of structured noise that becomes a problematically large confounding factor. When recordings are available from multiple electrodes over multiple time points, a linear multivariate filter can combine the signal from those electrodes and time points in order to actively counter interfering ERP components and other structured sources of noise.

The defining characteristic of the beamformer approach is the use of a template of the COI, which is both its biggest strength and weakness. The template

allows a researcher to exert strict control over the signal that is isolated by the filter, which is a desirable property then the focus of the study is a specific, well defined COI. By verifying that the template of the COI is not tainted by other ERP components and noise sources such as eye artifacts, the filter can effectively counter such noise sources. However, similar to the dependence of supervised learning techniques on the accuracy of the training labels, the performance of the beamformer is restricted by the method used to construct the template of the COI. For example, if the shape of the COI differs greatly between subjects and the method of estimating the template does not capture this (as is the case in the method we used), the beamformer will not perform optimally (figure 3.3, third column). Care must be taken to verify the validity of the template before drawing any conclusions about the output of the filter. For example, our method of constructing the template is only appropriate for isolating components that are known to have a relatively stable timing and scalp distribution, such as the N400 component. Note that the temporal pattern of the COI depends on the reference used in the recording. Therefore, the reference assumed by the template should match the reference used by the recording to which the beamformer filter is applied.

To be able to isolate components such as the N/P150, which exhibit polarity inversion [58], other approaches must be explored, such as using the summation of two beamformers, constructed using separate templates for the N150 and P150. Further data-driven elements could be added to the process, such as shifting the template in time to optimally fit the ERP, although the more data-driven elements are added, the more the method will behave like a supervised learning approach. In order to distinguish between the behavior of the beamformer approach in general and the behavior of the particular method we used to estimate the template, simulation results are provided in the supplementary information for a beamformer that uses the true activation pattern of the COI as a template.

When we wish to study the effect of some experimental manipulation on a specific COI, the results favor the beamformer over the supervised learning approaches. In this case, the specificity criterion mentioned in section 3.2.2 comes into play and it is not only important to have a good representation of the COI, but also to properly counter the influence of other ERP components. The simulation study shows that while the chained-LCMV beamformer is more robust against unstructured noise, the stLCMV beamformer was better able to counteract structured noise, which is more important in this case.

3.6 Conclusion

We have demonstrated the need for multivariate analysis, when ERP component amplitude measurements are desired on a single-trial basis.

The LCMV beamformer can be expanded from its traditional formulation as a spatial filter, to be a spatio-temporal filter which workings can be easily understood. The behavior of the beamformer, along with various supervised learning approaches, has been evaluated on simulated and real EEG data. Supervised

learning approaches, such as the lSVM, extract any information that aids in reproducing the labels of a training set, regardless of whether this information comes from the ERP component the researcher wishes to study (the COI), or from some other ERP component or from structured noise sources, such as EOG artifacts.

In contrast, the beamformer filter only isolates signals that conform to a given template. By crafting a template for the ERP component, which amplitude the researches wishes to measure, the beamformer effectively suppresses structured noise sources, such as other ERP components, eye-artifacts, and so forth. A simple method of crafting such a template, based on the grand average ERP of multiple subjects, is demonstrated to perform well. Of the two approaches used to extend the LCMV beamformer to the spatio-temporal domain, the stLCMV approach scored best during the evaluation. During the simulation study, the stLCMV beamformer output was mostly unaffected by the presence of interfering ERP components under all tested circumstances.

This means that if a researcher can construct a template that is a good approximation of the shape of the COI, the stLCMV beamformer output is mostly attributable to changes in its amplitude and is mostly unaffected by structured noise, such as other ERP components.

Supplementary information

Complete simulation study

The complete list of methods of estimating COI amplitude that were evaluated is:

ROI-mean Traditional method of taking the mean voltage over a few selected channels and time points (channels FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s).

perfect-chained-LCMV A chained-LCMV beamformer filter, which uses a separate spatial and temporal beamformer. The spatial and temporal templates were set to the true shape of the COI. Estimation of the covariance matrix was performed using OAS [158].

chained-LCMV A chained-LCMV beamformer filter, which uses a separate spatial and temporal beamformer. The spatial and temporal templates were constructed using the procedure outlined in section 3.3.5. Estimation of the covariance matrix was performed using OAS [158].

perfect-stLCMV An stLCMV beamformer filter, where the template of the COI was set to the true shape of the COI. Estimation of the covariance matrix was performed using OAS [158].

stLCMV An stLCMV beamformer filter, where the template of the COI was constructed using the procedure outlined in section 3.3.5. Estimation of the covariance matrix was performed using OAS [158].

ss-stLCMV An stLCMV beamformer filter, where the template of the COI was constructed using the procedure outlined in section 3.3.5, but limited to the training data of a single recording (the one currently being processed). Estimation of the covariance matrix was performed using OAS [158].

lSVM A linear support vector machine.

shrinkage-LDA Linear discriminant analysis, with shrinkage of the covariance matrix. Estimation of the shrunk covariance matrix was performed using OAS [158].

LR Logistic regression. Estimation for the amount of $\ell 2$ regularization was performed using OAS [158].

restricted-lSVM An lSVM that does not use all the features. Instead, it only uses a few selected channels and time points (channels FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s).

group-lSVM An lSVM that was trained on the pooled training data of all 10 subjects.

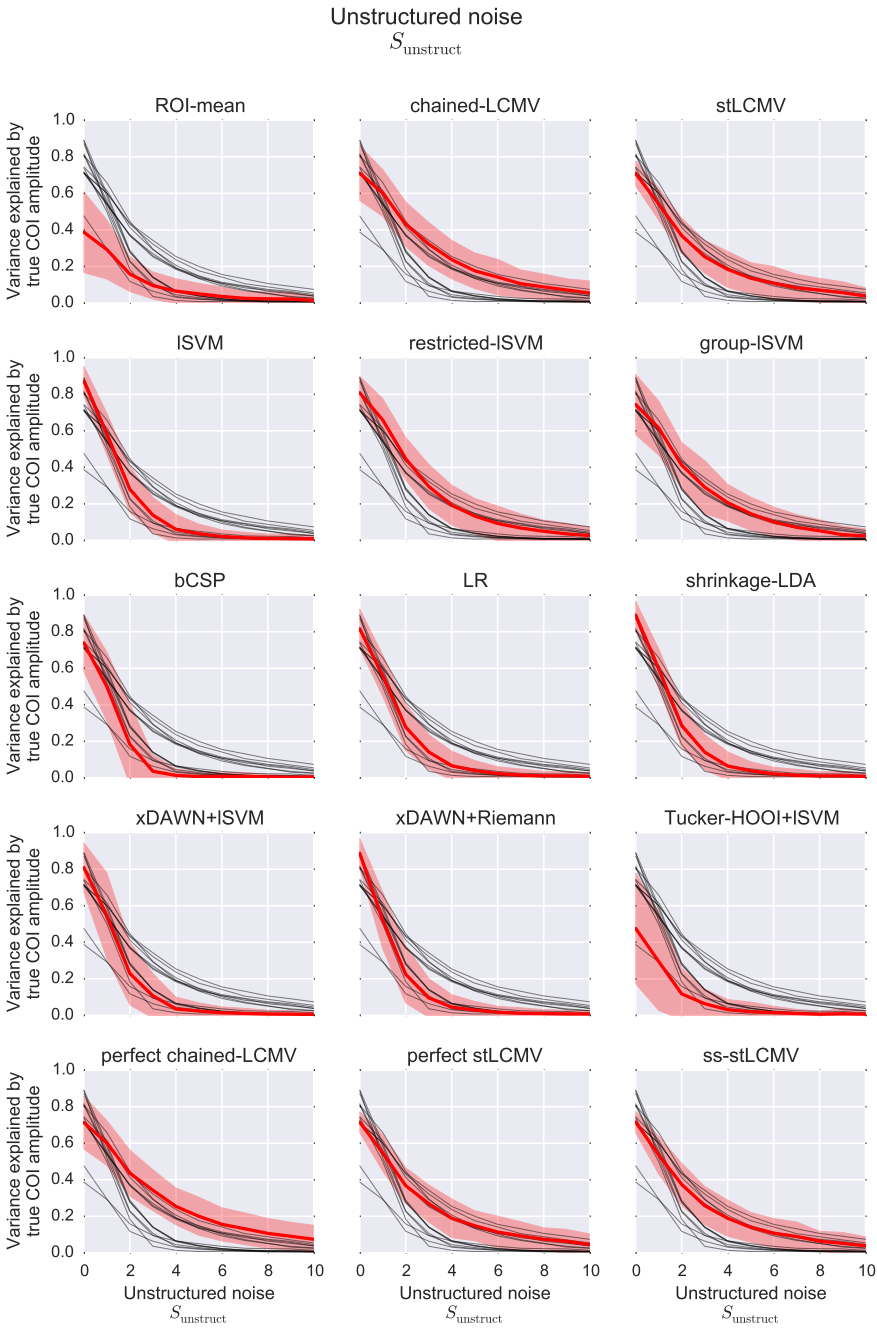
xDAWN+lSVM An xDAWN spatial filter [116], combined with an lSVM. The 4 most descriptive spatial components computed by the xDAWN algorithm were retained. Estimation of the covariance matrix was performed using OAS [158].

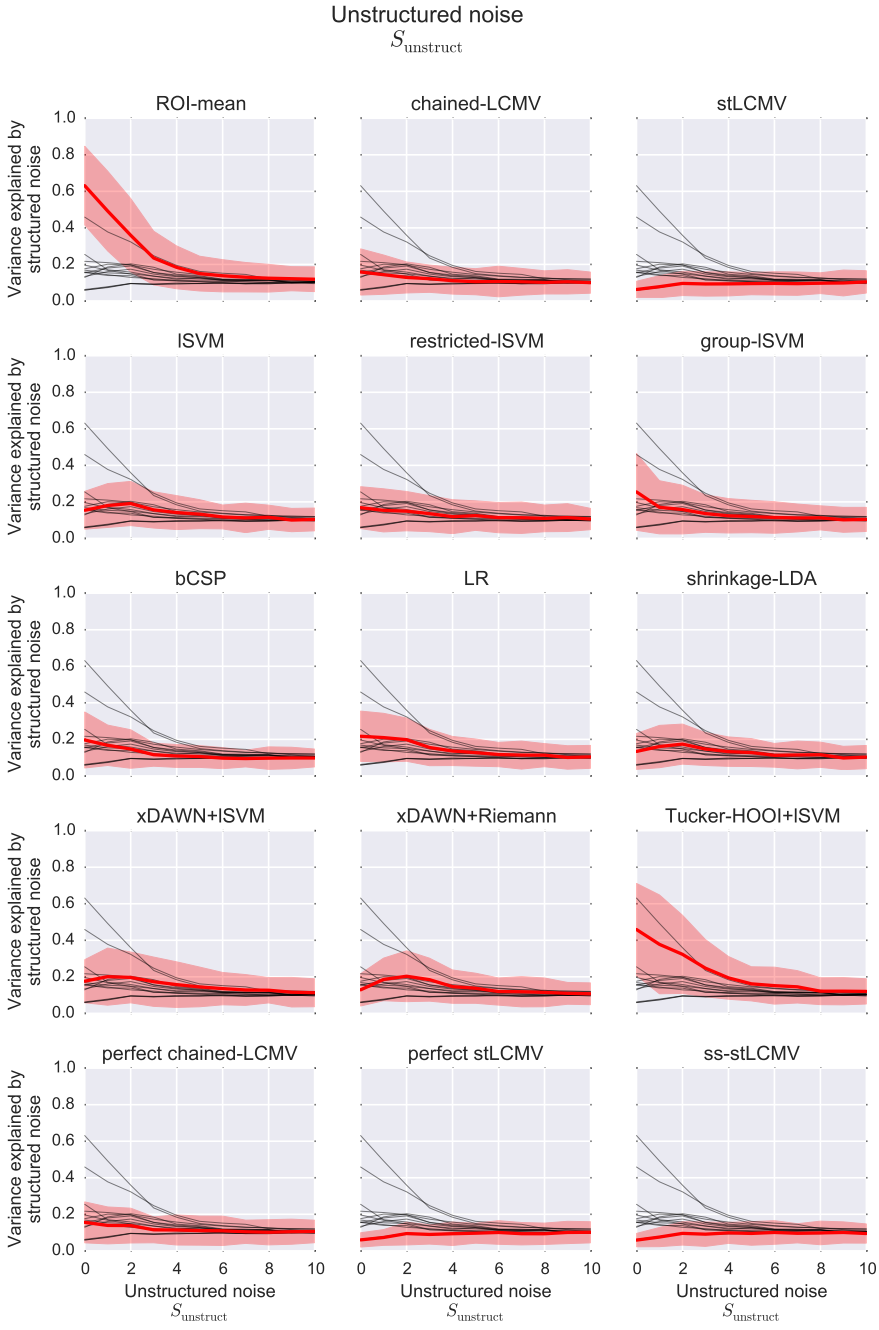
xDAWN+Riemann An xDAWN spatial filter [116], combined with classification of the covariance matrix on a Riemannian manifold [166]. The 4 most descriptive spatial components computed by the xDAWN algorithm were retained. Estimation of the covariance matrix was performed using OAS [158]. This is based on the approach that won the Kaggle DecMeg 2014 competition (https://www.kaggle.com/c/decoding-the-human-brain).

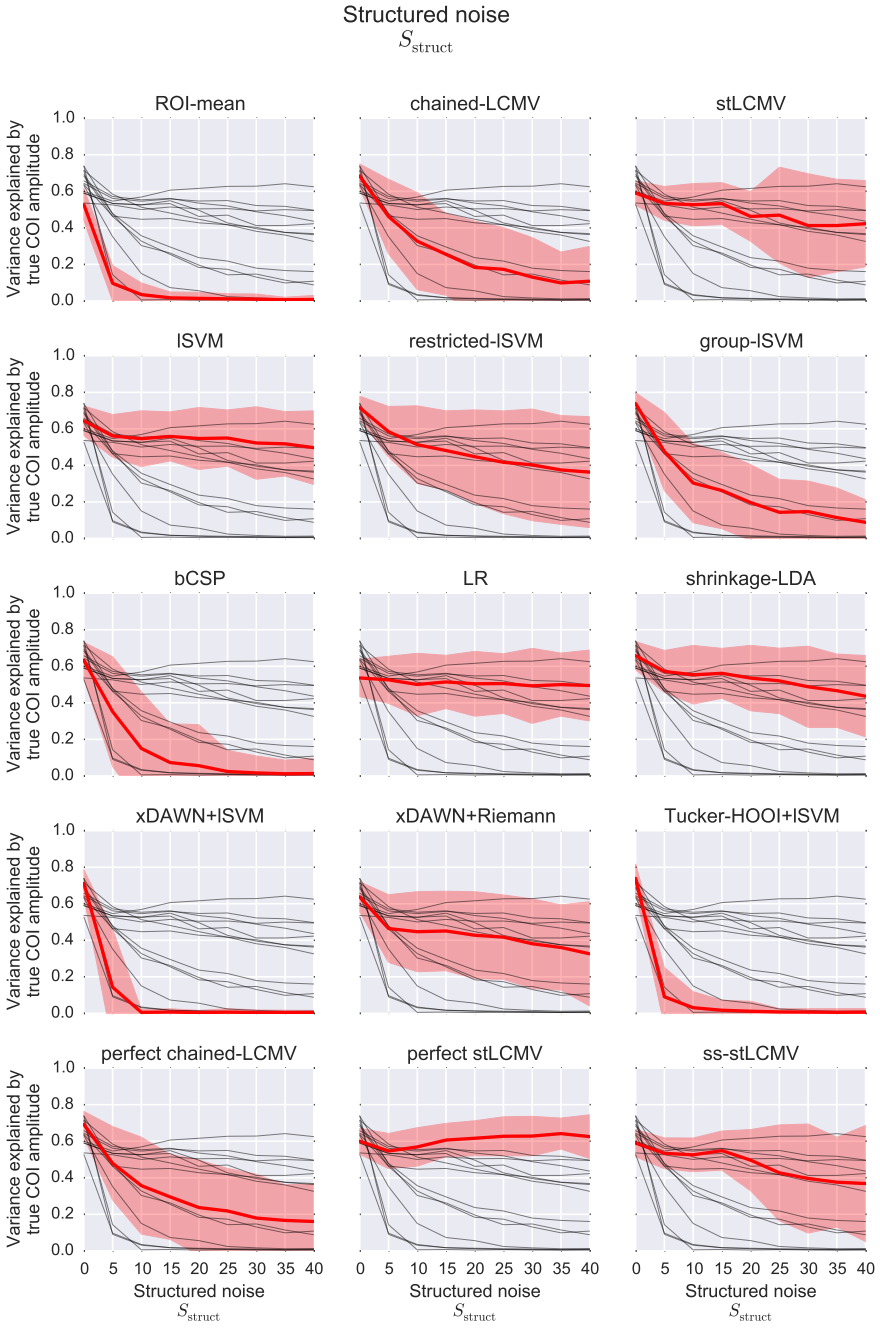
b**CSP** The bi-linear common spatial patterns algorithm [118].

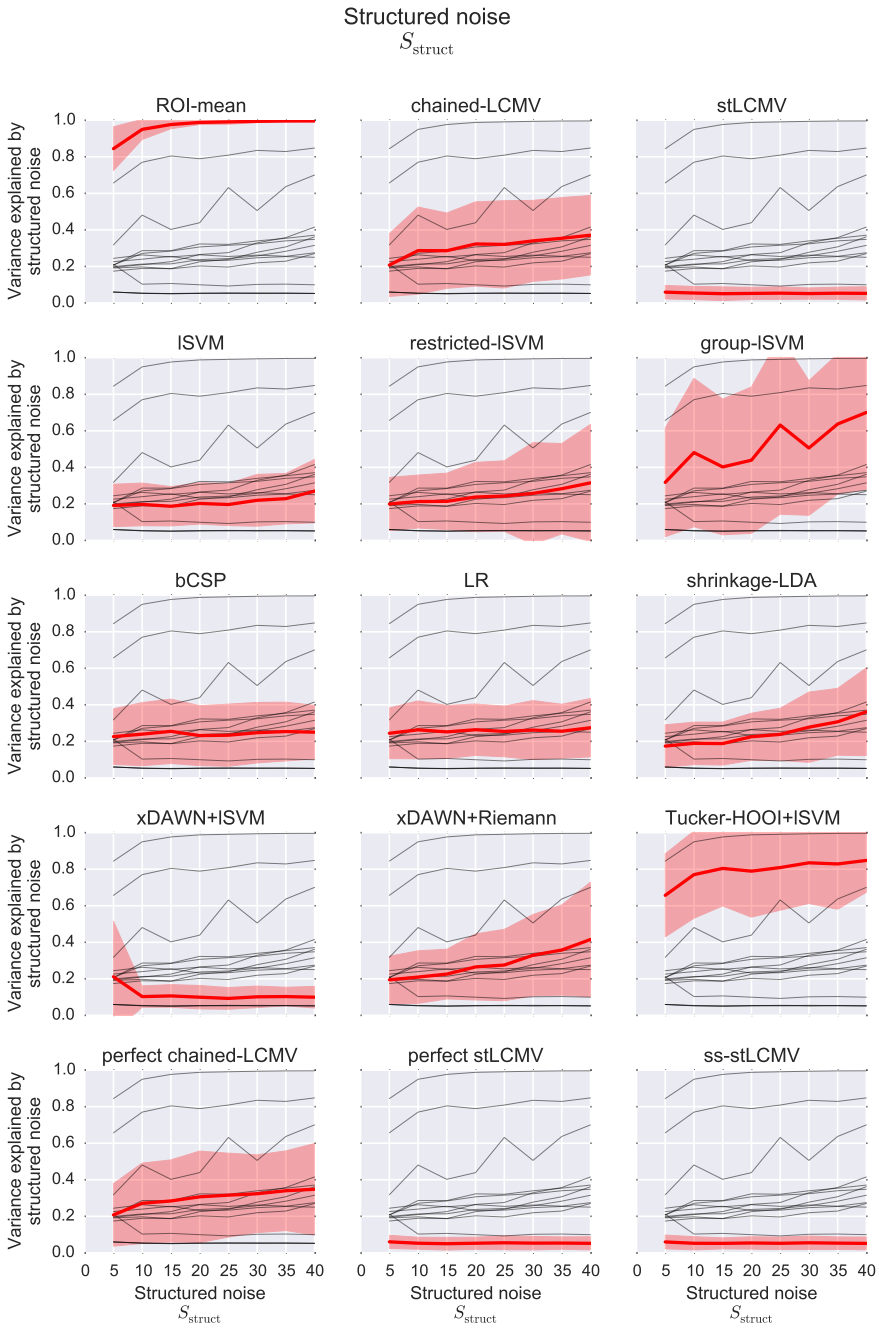
Tucker-HOOI+lSVM A combination of unsupervised tensor decomposition, using the Tucker-higher-order orthogonal iteration (HOOI) algorithm, and an lSVM. The tensor decomposition was used to map for each trial, the original (#channels × #samples ×) tensor onto a (8×10) core tensor, which was used as features for the lSVM.

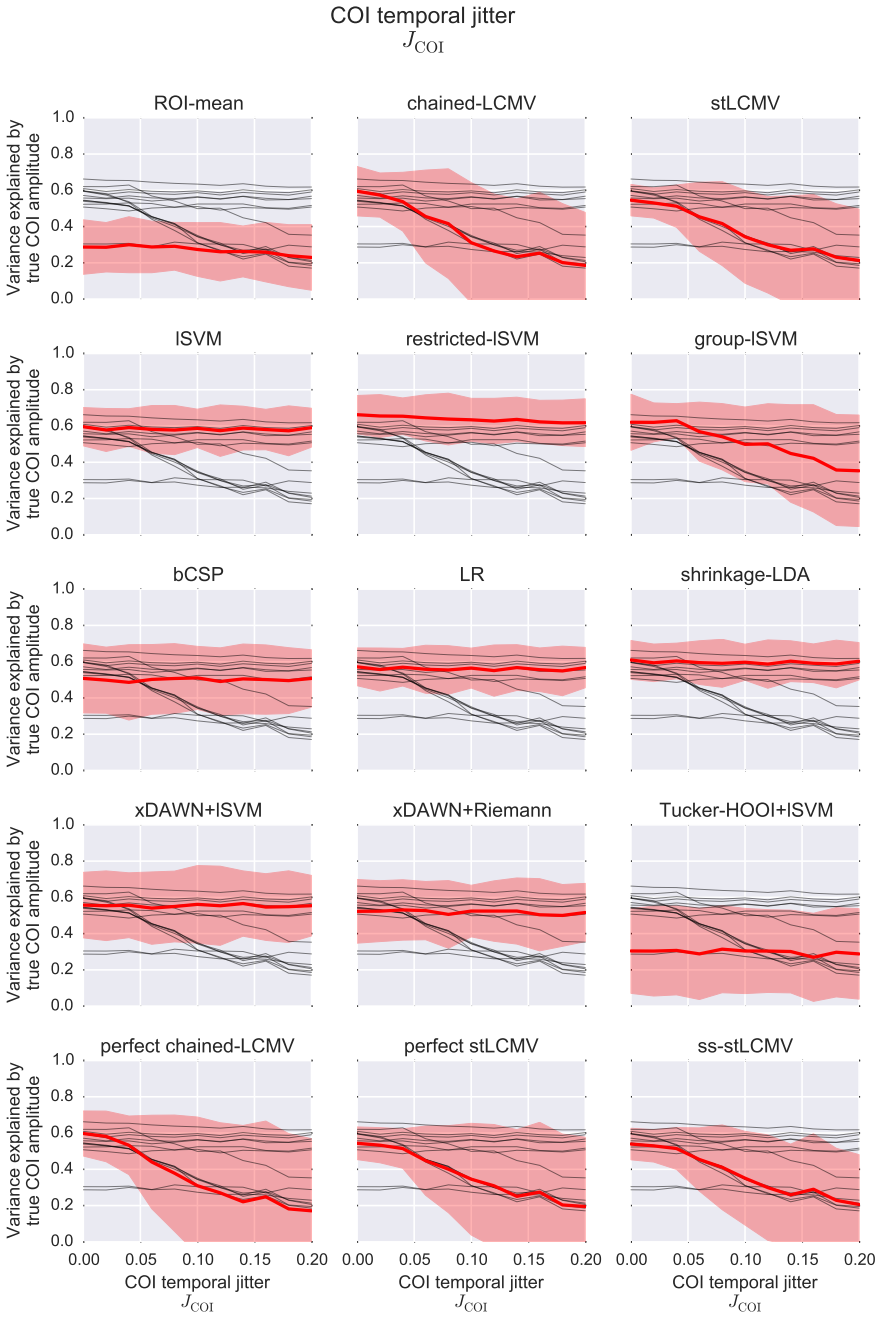
The following graphs show the performance of above techniques to estimate COI amplitude under various simulated conditions. The performance is quantified by two metrics: the variance of the output explained (R^2 stat) by the true amplitude of theCOI (higher is better) and the variance explained by least squares regression with the structured noise sources (lower is better). Each curve is the average of 100 runs. The 95% confidence interval is plotted as a shaded area. In each graph, the curve corresponding to a particular method is highlighted in red, with the curves corresponding to the other methods shown as thin black lines for comparison.

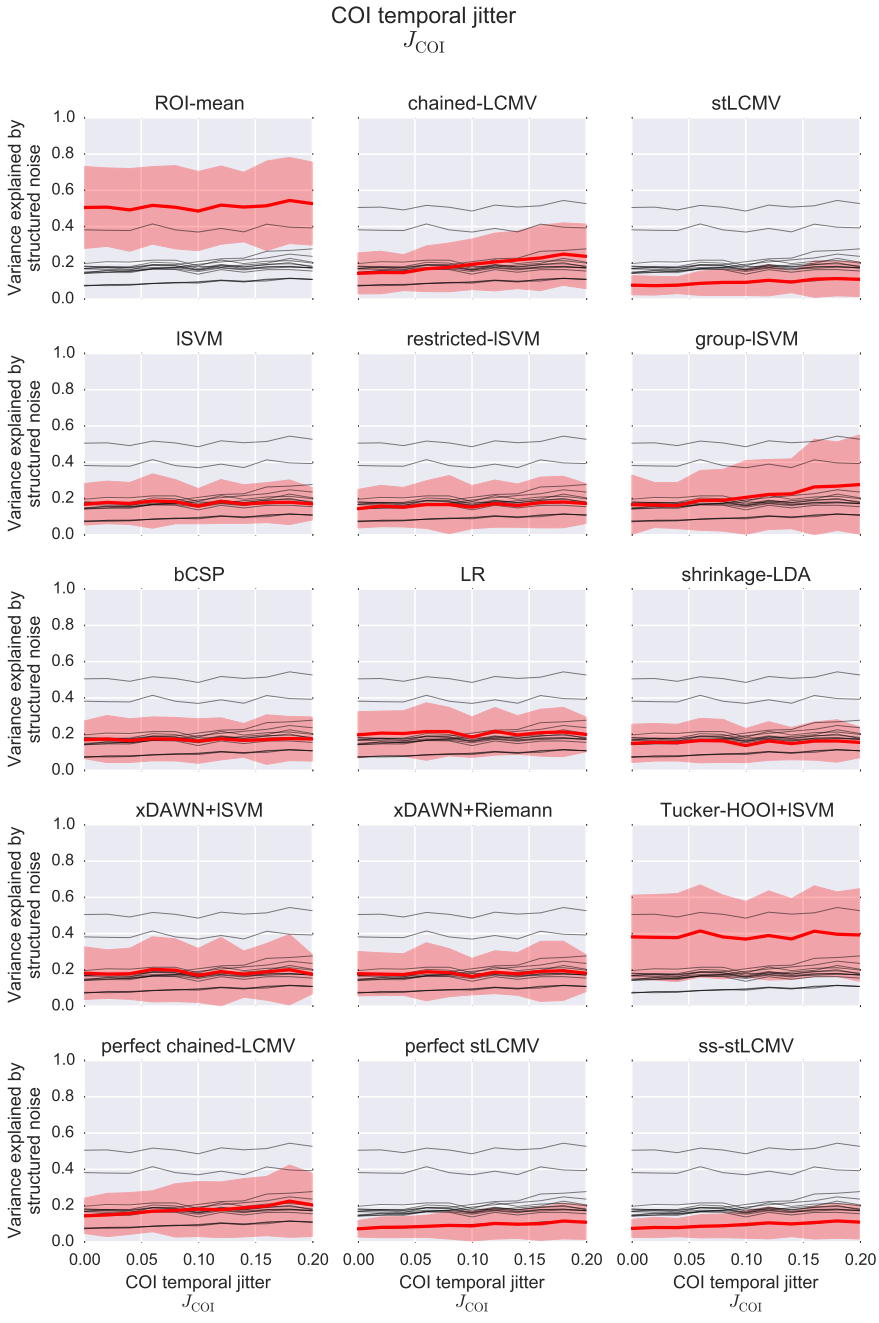

Unstructured noise
S_{unstruct}
ROI-mean
chained-LCMV
stLCMV
ISVM
restricted-ISVM
group-ISVM
bCSP
LR
shrinkage-LDA
xDAWN+ISVM
xDAWN+Riemann
Tucker-HOOI+ISVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Unstructured noise
S_{unstruct}

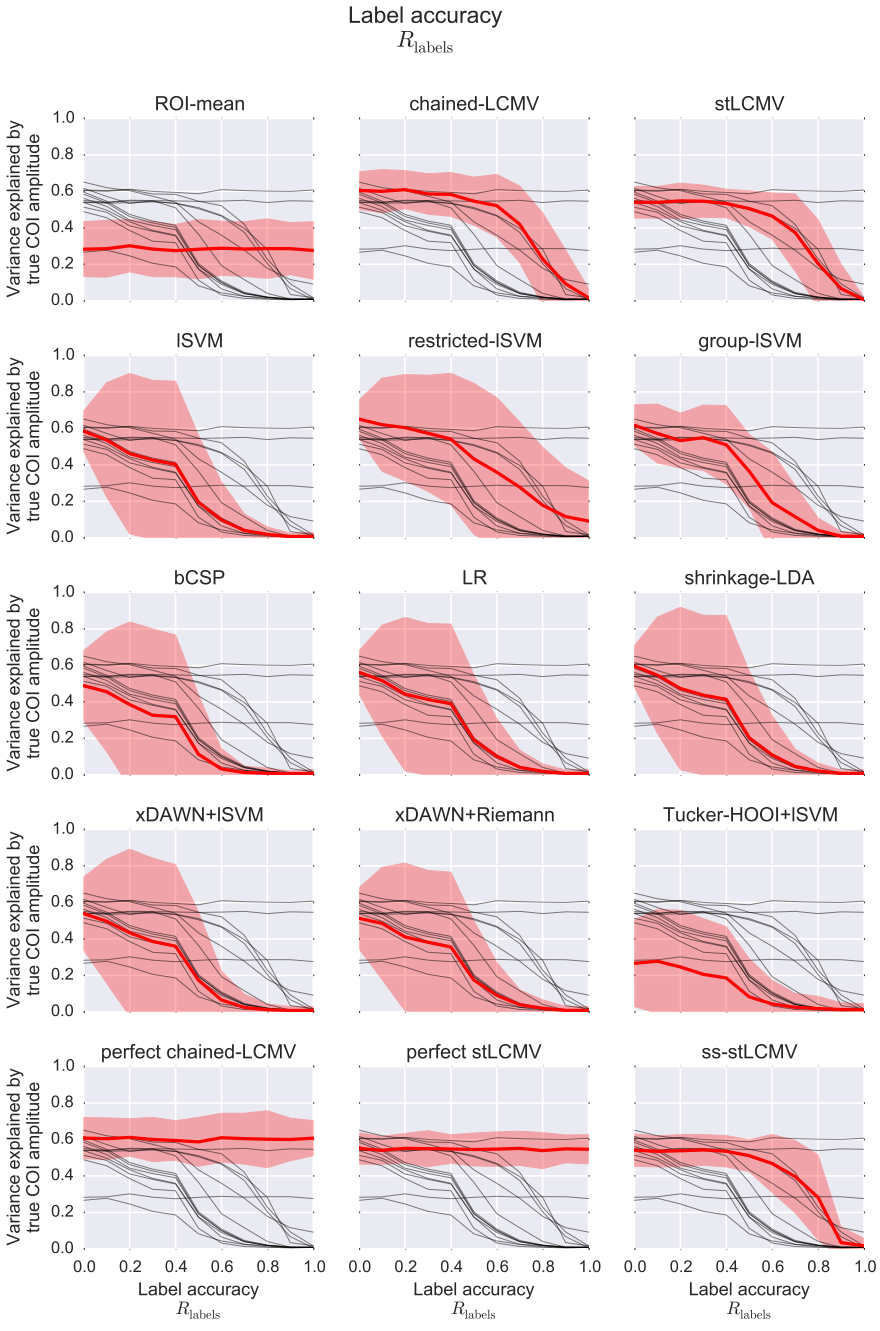

Unstructured noise
S_{unstruct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Unstructured noise
S_{unstruct}

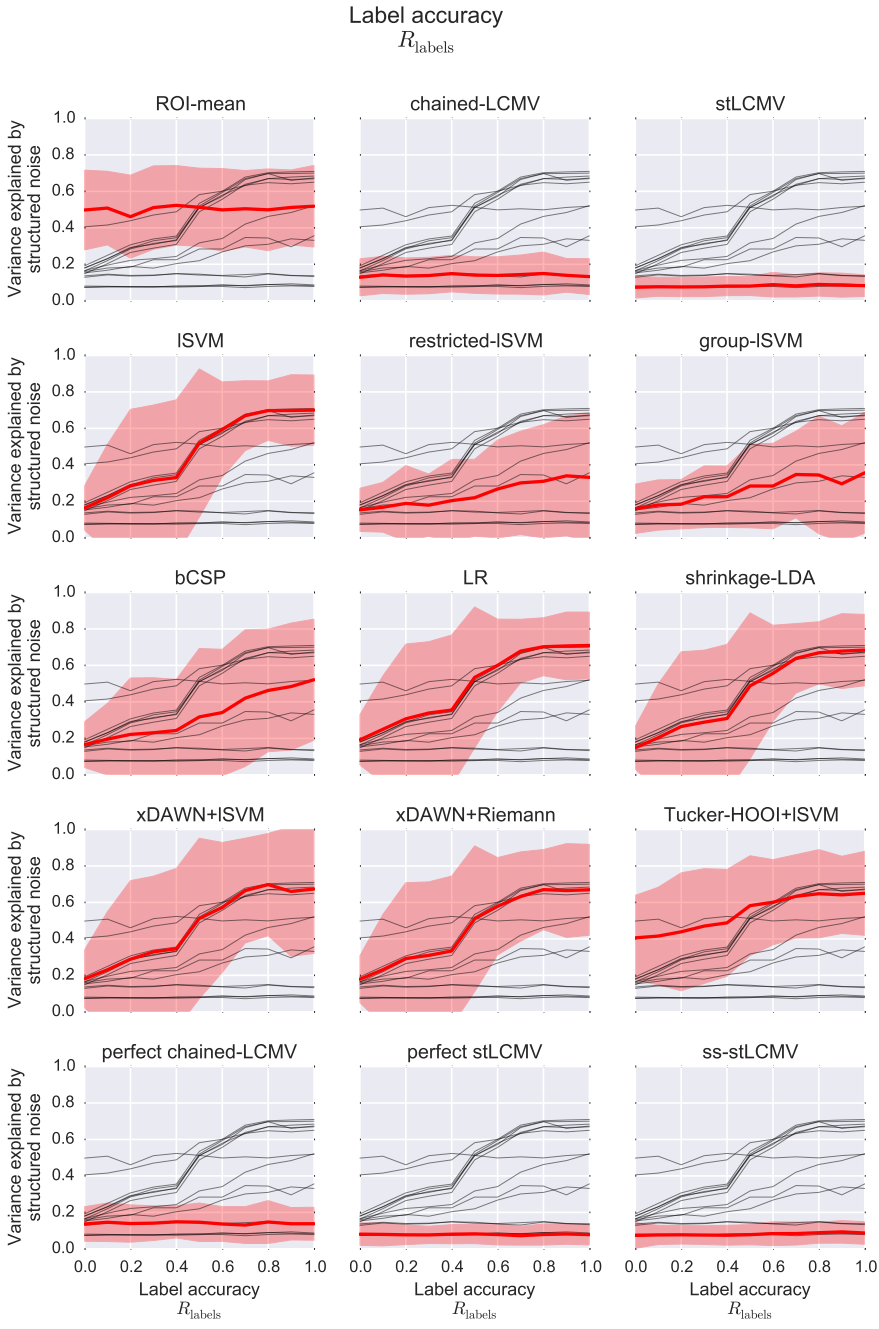

Structured noise
S_{struct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0 5 10 15 20 25 30 35 40
Structured noise
S_{struct}


Structured noise
S_{struct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0 5 10 15 20 25 30 35 40
Structured noise
S_{struct}


COI temporal jitter
J_{COI}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
COI temporal jitter
J_{COI}

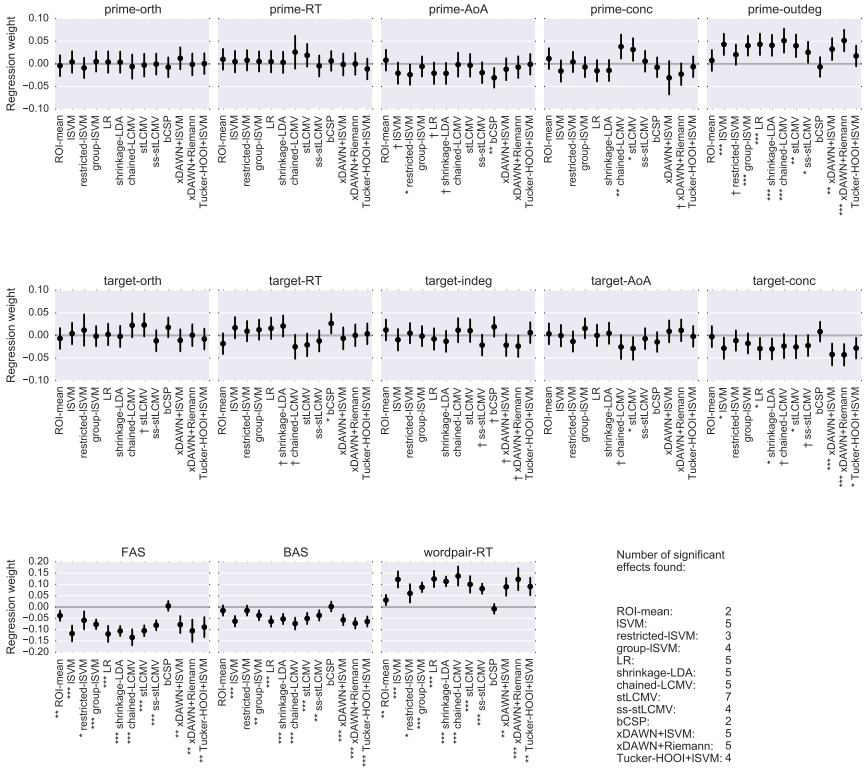

COI temporal jitter
J_{COI}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
COI temporal jitter
J_{COI}


Label accuracy
R_{labels}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
Label accuracy
R_{labels}


Label accuracy
R_{labels}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
Label accuracy
R_{labels}

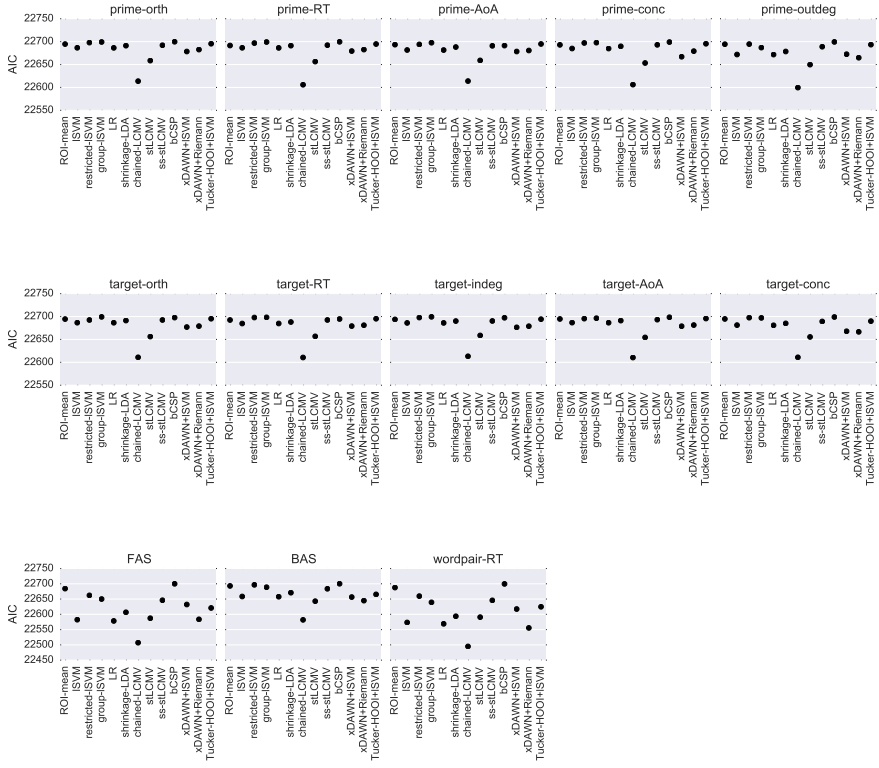
Evaluation on real EEG data

Univariate regression between each analysis method and each stimulus property. Dots indicate the regression weights obtained through the LME model and vertical lines indicate the 95% confidence interval. *P*-values are given for the null-hypothesis that the regression coefficient is zero. All variables were *z*-transformed before being entered into the model, hence regression weights can be interpreted as Pearson's correlation coefficients.



† p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001

Akaike information criterion (AIC) measurements for the regression models used in the previous figure.



The following tables contain the raw values obtained during the regression study on real EEG data. Given are the regression coefficients (coeff), confidence interval of these coefficients (low-hight), t-values, p-values, and the AIC.

N400 amplitude estimation through ROI-mean

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.004040	-0.027168	0.019088	-0.342396	0.73214	22694.635799
prime-RT	0.009939	-0.013180	0.033058	0.842584	0.39971	22691.466089
prime-AoA	0.008094	-0.015029	0.031217	0.686067	0.49287	22693.355484
prime-conc	0.011547	-0.011568	0.034663	0.979078	0.32784	22692.909825
prime-outdeg	0.007461	-0.015663	0.030584	0.632353	0.52734	22694.353255
target-orth	-0.006705	-0.029830	0.016420	-0.568268	0.57001	22694.430162
target-RT	-0.018138	-0.041234	0.004957	-1.539292	0.12413	22692.387113
target-indeg	0.012063	-0.011052	0.035177	1.022847	0.30669	22693.707495
target-AoA	0.003501	-0.019627	0.026629	0.296705	0.76677	22694.664996
target-conc	-0.002645	-0.025774	0.020484	-0.224133	0.82271	22694.701653
FAS	-0.038215	-0.061192	-0.015237	-3.259705	0.00116	22684.197385
BAS	-0.015554	-0.038659	0.007550	-1.319473	0.18739	22693.013910
wordpair-RT	0.030776	0.007745	0.053807	2.619115	0.00898	22687.635866

N400 amplitude estimation through lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.004132	-0.019706	0.027970	0.339735	0.73415	22686.536836
prime-RT	0.004865	-0.018972	0.028702	0.400012	0.68925	22686.492254
prime-AoA	-0.020635	-0.044432	0.003162	-1.699565	0.0896	22681.683334
prime-conc	-0.015865	-0.039680	0.007949	-1.305746	0.19201	22684.949092
prime-outdeg	0.042976	0.019324	0.066629	3.561209	0.000390993	22671.824437
target-orth	0.004167	-0.019671	0.028005	0.342628	0.73197	22686.534863
target-RT	0.016951	-0.006859	0.040762	1.395335	0.1633	22684.707656
target-indeg	-0.009654	-0.033484	0.014176	-0.794003	0.42743	22686.022056
target-AoA	-0.000406	-0.024246	0.023434	-0.033377	0.97338	22686.651134
target-conc	-0.028301	-0.052060	-0.004542	-2.334664	0.01981	22681.220098
FAS	-0.117075	-0.152427	-0.081723	-6.490809	8.93444×10^{-5}	22582.440434
BAS	-0.062610	-0.086051	-0.039169	-5.234984	2.11096×10^{-7}	22658.404367
wordpair-RT	0.121639	0.085557	0.157722	6.607303	8.13534×10^{-5}	22573.382098

N400 amplitude estimation through restricted-lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.009039	-0.031624	0.013546	-0.784387	0.43305	22697.848354
prime-RT	0.007834	-0.014753	0.030421	0.679804	0.49683	22696.755648
prime-AoA	-0.023880	-0.046413	-0.001347	-2.077116	0.03811	22693.882183
prime-conc	0.004155	-0.018437	0.026747	0.360448	0.71861	22697.057827
prime-outdeg	0.020282	-0.002268	0.042832	1.762850	0.07831	22694.551178
target-orth	0.011686	-0.023001	0.046373	0.660311	0.52473	22692.426463
target-RT	0.009407	-0.013177	0.031992	0.816417	0.41451	22697.797122
target-indeg	0.004980	-0.017611	0.027571	0.432092	0.66579	22697.742386
target-AoA	-0.013379	-0.035954	0.009195	-1.161605	0.24574	22695.360678
target-conc	-0.011426	-0.034006	0.011154	-0.991778	0.32161	22697.480361
FAS	-0.058786	-0.098656	-0.018915	-2.889806	0.0175	22662.072540
BAS	-0.015827	-0.038394	0.006740	-1.374601	0.16964	22696.576084
wordpair-RT	0.060365	0.019553	0.101177	2.898960	0.01727	22659.583732

N400 amplitude estimation through group-lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.005026	-0.017347	0.027399	0.440312	0.65983	22699.181945
prime-RT	0.005237	-0.017136	0.027610	0.458803	0.6465	22699.116376
prime-AoA	-0.005855	-0.028227	0.016517	-0.512953	0.60813	22697.583772
prime-conc	-0.007159	-0.029529	0.015211	-0.627224	0.53069	22697.748803
prime-outdeg	0.040133	0.017931	0.062335	3.542895	0.000 418 72	22686.921270
target-orth	-0.001627	-0.024002	0.020749	-0.142480	0.88674	22699.199613
target-RT	0.012554	-0.009805	0.034913	1.100496	0.27145	22698.165623
target-indeg	-0.001631	-0.024007	0.020744	-0.142885	0.88642	22699.337064
target-AoA	0.015583	-0.006766	0.037933	1.366597	0.17214	22696.285612
target-conc	-0.017602	-0.039944	0.004741	-1.544086	0.12296	22696.995146
FAS	-0.076316	-0.098167	-0.054466	-6.845464	8.1859×10^{-12}	22650.034361
BAS	-0.036827	-0.059057	-0.014597	-3.247006	0.00121	22688.729961
wordpair-RT	0.087092	0.065260	0.108923	7.818924	5.9952×10^{-15}	22639.256399

N400 amplitude estimation through LR

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.003946	-0.019906	0.027798	0.324227	0.74585	22686.354425
prime-RT	0.004862	-0.018989	0.028713	0.399547	0.6896	22686.299919
prime-AoA	-0.020730	-0.044540	0.003080	-1.706437	0.08832	22681.570630
prime-conc	-0.015333	-0.039162	0.008497	-1.261078	0.20765	22684.870805
prime-outdeg	0.043071	0.019405	0.066737	3.567061	0.000382502	22671.719503
target-orth	0.002228	-0.021625	0.026081	0.183069	0.85479	22686.426027
target-RT	0.015814	-0.008014	0.039642	1.300778	0.19371	22684.769306
target-indeg	-0.007924	-0.031772	0.015923	-0.651301	0.51504	22686.035460
target-AoA	-0.000009	-0.023862	0.023845	-0.000713	0.99943	22686.459541
target-conc	-0.028891	-0.052660	-0.005122	-2.382303	0.01744	22680.804234
FAS	-0.119048	-0.154880	-0.083215	-6.511703	8.85733×10^{-5}	22578.378878
BAS	-0.063613	-0.087056	-0.040171	-5.318630	1.35836×10^{-7}	22657.666512
wordpair-RT	0.123920	0.087866	0.159974	6.736539	7.08355×10^{-5}	22568.987026

N400 amplitude estimation through shrinkage-LDA

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.003632	-0.019844	0.027109	0.303241	0.76179	22691.136628
prime-RT	0.003284	-0.020193	0.026761	0.274141	0.78405	22691.153547
prime-AoA	-0.021084	-0.044517	0.002348	-1.763547	0.07819	22688.124636
prime-conc	-0.014632	-0.038088	0.008824	-1.222655	0.22182	22689.733660
prime-outdeg	0.040924	0.017618	0.064230	3.441589	0.000608349	22678.435013
target-orth	-0.002185	-0.025662	0.021293	-0.182390	0.85532	22691.195431
target-RT	0.020786	-0.002648	0.044219	1.738484	0.08251	22687.984372
target-indeg	-0.012939	-0.036400	0.010522	-1.080946	0.28005	22690.061106
target-AoA	0.004831	-0.018644	0.028307	0.403352	0.6868	22691.066020
target-conc	-0.029850	-0.053236	-0.006463	-2.501623	0.01256	22684.994958
FAS	-0.105816	-0.128118	-0.083514	-9.299325	0	22606.672059
BAS	-0.052891	-0.076081	-0.029701	-4.470262	8.94185×10^{-6}	22670.815820
wordpair-RT	0.112956	0.090823	0.135089	10.002662	0	22593.667117

N400 amplitude estimation through chained-LCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.006158	-0.032971	0.020654	-0.450170	0.65271	22613.682123
prime-RT	0.025763	-0.010770	0.062296	1.382155	0.18875	22605.842816
prime-AoA	-0.001758	-0.028574	0.025058	-0.128513	0.89778	22613.877213
prime-conc	0.038022	0.011336	0.064709	2.792519	0.00535	22606.083472
prime-outdeg	0.051086	0.024505	0.077667	3.766800	0.000 177 495	22599.625823
target-orth	0.022466	-0.004305	0.049237	1.644807	0.1004	22610.953388
target-RT	-0.025026	-0.051786	0.001734	-1.832972	0.06718	22610.576280
target-indeg	0.011587	-0.015217	0.038392	0.847285	0.39709	22613.211453
target-AoA	-0.025434	-0.052192	0.001325	-1.862948	0.06284	22610.460809
target-conc	-0.023575	-0.050342	0.003191	-1.726286	0.08468	22610.954501
FAS	-0.133987	-0.169488	-0.098486	-7.397234	$7.564\,65 \times 10^{-6}$	22507.132270
BAS	-0.073358	-0.099687	-0.047029	-5.460765	$6.333\,12 \times 10^{-8}$	22581.880911
wordpair-RT	0.136897	0.094242	0.179553	6.290252	$5.865\,31 \times 10^{-5}$	22494.922880

N400 amplitude estimation through stLCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.002728	-0.028002	0.022546	-0.211541	0.83252	22658.543987
prime-RT	0.018824	-0.006417	0.044065	1.461664	0.14423	22656.384533
prime-AoA	-0.003273	-0.028547	0.022000	-0.253844	0.79968	22658.971728
prime-conc	0.031654	0.006475	0.056834	2.463963	0.01395	22653.304790
prime-outdeg	0.039872	0.014749	0.064995	3.110584	0.00193	22649.735339
target-orth	0.022895	-0.002330	0.048120	1.778916	0.07563	22656.194661
target-RT	-0.020971	-0.046204	0.004262	-1.628890	0.10373	22656.704066
target-indeg	0.010416	-0.014849	0.035680	0.808046	0.4193	22658.700280
target-AoA	-0.028793	-0.053989	-0.003598	-2.239816	0.02538	22654.351862
target-conc	-0.025503	-0.050716	-0.000290	-1.982528	0.04776	22655.430540
FAS	-0.104778	-0.128986	-0.080570	-8.483133	0	22587.230850
BAS	-0.050568	-0.075598	-0.025537	-3.959599	$8.177\,16 \times 10^{-5}$	22642.986334
wordpair-RT	0.099884	0.062942	0.136827	5.299295	0.000 256 578	22590.774994

N400 amplitude estimation through ss-stLCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.000663	-0.023926	0.022599	-0.055880	0.95545	22692.034744
prime-RT	-0.004635	-0.027896	0.018626	-0.390552	0.69623	22692.210717
prime-AoA	-0.019508	-0.042732	0.003715	-1.646425	0.10007	22690.802730
prime-conc	0.005807	-0.017452	0.029067	0.489370	0.62471	22692.971923
prime-outdeg	0.025509	0.002313	0.048704	2.155442	0.03143	22688.876381
target-orth	-0.011864	-0.035112	0.011384	-1.000221	0.31751	22692.442320
target-RT	-0.012153	-0.035400	0.011095	-1.024591	0.30587	22692.459760
target-indeg	-0.021352	-0.044568	0.001863	-1.802645	0.07182	22690.265920
target-AoA	-0.007080	-0.030338	0.016177	-0.596686	0.55089	22693.152902
target-conc	-0.022295	-0.045506	0.000916	-1.882601	0.06012	22689.194723
FAS	-0.080526	-0.103108	-0.057943	-6.988844	5.84865×10^{-12}	22646.099686
BAS	-0.037115	-0.060235	-0.013995	-3.146358	0.00171	22683.670117
wordpair-RT	0.080815	0.058237	0.103393	7.015503	4.88853×10^{-12}	22645.747919

N400 amplitude estimation through bCSP

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.007585	-0.029499	0.014328	-0.678431	0.49752	22699.699278
prime-RT	0.006358	-0.015556	0.028272	0.568646	0.56961	22699.555197
prime-AoA	-0.030699	-0.052603	-0.008795	-2.746894	0.00603	22691.344031
prime-conc	-0.007626	-0.029539	0.014288	-0.682031	0.49524	22699.010468
prime-outdeg	-0.006883	-0.028797	0.015031	-0.615606	0.53817	22699.358077
target-orth	0.017754	-0.004157	0.039665	1.588113	0.1123	22697.637828
target-RT	0.026399	0.004492	0.048305	2.361833	0.01821	22694.583221
target-indeg	0.018994	-0.002916	0.040905	1.699089	0.08934	22697.273149
target-AoA	-0.014552	-0.036464	0.007360	-1.301603	0.19309	22698.378897
target-conc	0.008427	-0.013487	0.030340	0.753678	0.45106	22698.923887
FAS	0.004726	-0.017188	0.026641	0.422720	0.67251	22699.980843
BAS	0.002099	-0.019815	0.024013	0.187734	0.85109	22700.124289
wordpair-RT	-0.008095	-0.030008	0.013819	-0.723988	0.46909	22699.618566

N400 amplitude estimation through xDAWN+lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.012187	-0.012104	0.036479	0.983325	0.32575	22678.403599
prime-RT	-0.001437	-0.025743	0.022870	-0.115850	0.9078	22679.356521
prime-AoA	-0.011940	-0.036233	0.012352	-0.963364	0.33566	22678.412169
prime-conc	-0.030780	-0.067748	0.006189	-1.631865	0.13156	22667.003945
prime-outdeg	0.032481	0.008279	0.056683	2.630406	0.00869	22672.480694
target-orth	-0.010846	-0.035141	0.013449	-0.874997	0.38184	22676.994454
target-RT	-0.006418	-0.030720	0.017884	-0.517607	0.60488	22679.102071
target-indeg	-0.021390	-0.045651	0.002871	-1.727991	0.08438	22676.389555
target-AoA	0.009068	-0.015230	0.033366	0.731442	0.46472	22678.834819
target-conc	-0.042019	-0.066150	-0.017888	-3.412821	0.000675295	22667.802775
FAS	-0.077537	-0.115237	-0.039836	-4.030950	0.00225	22632.140305
BAS	-0.056740	-0.080726	-0.032754	-4.636414	4.13978×10^{-6}	22656.491877
wordpair-RT	0.088882	0.049475	0.128289	4.420712	0.00127	22617.324799

N400 amplitude estimation through xDAWN+Riemann

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.001219	-0.025330	0.022892	-0.099084	0.9211	22682.589326
prime-RT	-0.000110	-0.024222	0.024001	-0.008954	0.99286	22682.599063
prime-AoA	-0.007310	-0.031416	0.016796	-0.594325	0.55246	22680.589373
prime-conc	-0.022737	-0.046797	0.001323	-1.852221	0.06436	22679.175766
prime-outdeg	0.051539	0.027694	0.075384	4.236272	2.53731×10^{-5}	22664.851713
target-orth	0.000465	-0.023646	0.024577	0.037810	0.96985	22679.022175
target-RT	0.000300	-0.023811	0.024412	0.024396	0.98054	22680.985276
target-indeg	-0.023649	-0.047705	0.000407	-1.926840	0.05435	22678.849632
target-AoA	0.011183	-0.012916	0.035282	0.909535	0.36334	22681.322913
target-conc	-0.042766	-0.066695	-0.018837	-3.502924	0.000485746	22666.534226
FAS	-0.105084	-0.154679	-0.055490	-4.152887	0.00222	22583.658832
BAS	-0.072020	-0.095609	-0.048431	-5.984080	3.2833×10^{-9}	22644.821789
wordpair-RT	0.122593	0.074117	0.171069	4.956608	0.000704892	22555.531296

N400 amplitude estimation through Tucker-HOOI+lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.000494	-0.022531	0.023520	0.042091	0.96644	22695.366497
prime-RT	-0.011288	-0.034300	0.011724	-0.961424	0.33663	22694.715256
prime-AoA	-0.001047	-0.024072	0.021978	-0.089110	0.92902	22694.853662
prime-conc	-0.006462	-0.029483	0.016559	-0.550165	0.58236	22695.336435
prime-outdeg	0.017462	-0.005531	0.040456	1.488510	0.13701	22693.367150
target-orth	-0.007909	-0.030927	0.015110	-0.673400	0.50089	22695.185720
target-RT	0.003484	-0.019540	0.026508	0.296572	0.76687	22695.136349
target-indeg	0.006201	-0.016820	0.029222	0.527966	0.59767	22694.204781
target-AoA	-0.002191	-0.025215	0.020834	-0.186471	0.85212	22695.604288
target-conc	-0.028147	-0.051090	-0.005205	-2.404645	0.01641	22689.877562
FAS	-0.088982	-0.133312	-0.044653	-3.934208	0.00331	22620.853750
BAS	-0.064052	-0.086645	-0.041459	-5.556654	3.74762×10^{-8}	22665.335565
wordpair-RT	0.090631	0.051688	0.129573	4.561396	0.00129	22624.687438

Software used in this study

This study used the following software packages:

- Stimulus presentation was performed using MATLAB in combination with the Psychophysics toolbox [134].
- Dipole modeling and forward model computation was provided by the OpenMEEG package [163, 167].
- Data analysis was performed using Python in combination with the NumPy and SciPy packages [133].
- Plots were created using the Matplotlib package [134].
- The implementation of the lSVM, logistic regression (LR), shrinkage-LDA and OAS algorithms was provided by the Scikit-learn package [168].
- The implementation of the xDAWN algorithm and Riemannian distance calculations were provided by the pyRiemann package [169].
- The implementation of the Tucker-HOOI tensor decomposition algorithm was provided by the Scikit-tensor package [170].
- Statistical analysis was performed using R [171] in combination with the LME4 [172] and lmerTest [173] packages.

An implementation of the chained-LCMV and stLCMV beamformers used in this study, as well as methods to estimate a suitable template, can be found at: https://github.com/wmvanvliet/ERP-beamformer

Chapter 4

Using the N400 to cluster words into categories

4.1 Abstract

In this study, measurements of the N400 ERP component were utilized to cluster words that belong to the same semantic category. The starting point was an unordered list of 14 words which where exemplars of one of two possible semantic categories. By presenting each possible word-pair in a semantic priming setting and estimating the resulting N400 amplitude, the relationship strengths between the words was assessed. The words were then clustered into two groups, following the criterion that words belonging to the same group should evoke a small N400 amplitude when presented as a word-pair, and words belonging to different groups should evoke a large N400 amplitude. The result was a faithful reproduction of the original semantic categories: animals versus furniture items. This work represents a first step towards creating a semantic network that is based on measure of relationship strength that is derived from brain activity: the N400 component.

4.2 Introduction

Semantic priming studies [1] have demonstrated that our semantic machinery leverages relationships between words, indicating that not only the words themselves, but also relationships between those words are a part of our semantic memory [3, 21, 22]. Two frequently used measurements to quantify semantic priming are the RT of the subject in response to the target word [2] and the amplitude of the N400 component in the ERP [6, 7]. With these measurements, many types of relationships between words have been shown to affect semantic priming. For example, associative relationships [45, 47, 174], semantic category membership [6, 99, 102, 103], feature overlap [25], lexical relationships (e.g.rhymes [39]), taxonomic relations [175] (i.e."is a") and other types of world knowledge [84] (e.g."is used to", "works in", "is made of", etc.).

Given these relationships, a useful way to model our semantic memory is as a network, where the nodes represent words and the edges represent the relationships that use utilized during semantic processing [21]. The overall structure of the network and its behavior as new words and relationships are learned and forgotten can teach us much about the working of our semantic memory.

One way of mapping out a semantic network is to study patterns in the production of language. For example, in association norm studies [45, 47] thou-

sands of participants are asked to write down their first (or top three) associations to a given *cue* word. By using the notion of FAS, the number of participants that wrote down the *target* word in response to the cue, huge semantic networks can be mapped out [46]. A downside of free association studies is that participants are allowed to think a bit, before they write down their response (even if they are encouraged not to), making it a conscious decision. The true first associates to a cue word might be discarded if are socially unacceptable (curse words, racist beliefs, etc.) or if the participant cannot rationally explain why they came to mind. However, word relationships are utilized in the brain at an early stage, even before conscious awareness [28] and socially unacceptable relationships are present in our semantic memory [176].

A semantic network that is based on unconscious physiological responses might more accurately mirror the relationships represented in our semantic memory. The best responses might be those generated by the brain itself. This makes the amplitude of the N400 components a desirable measure as a basis for semantic networks. There are however, a few obstacles that make this challenging. Three of them are addressed in this study.

The first challenge is the inherent noise in an EEG recording. While proper experimental design (chapter 2) and advanced multivariate techniques (chapter 3) both improve the accuracy of the N400 amplitude estimates, there is still a long way to go before single-trial estimates are accurate enough to be useful as measure of relationship strength. In order to obtain a reliable measure, an average of multiple N400 estimates must be made, which leads to the second challenge.

The second challenge is the effect of stimulus-repetition on the semantic priming effect. The amplitude of the N400 component diminishes as words are re-used during the same session [177]. However, there seems to be a plateau where the amplitude of the N400 is not reduced any further, even when stimuli are massively repeated [99, 104], depending on the task given to the subject [105].

The third challenge is the fact that in order to construct a full semantic network with n words, without any prior assumptions, each of the $n \times (n-1)$ possible word-pairs must be presented to the subject. As the number of words grows, this number becomes problematically large and a decision must be made which word-pairs to present.

The current study is a first step towards tackling these three problems and towards a semantic network that uses the amplitude of the N400 component as measure of relationship strength.

A possible way to deal with exponential growth of possible word-pairs is to take an hierarchical approach. Earlier research has already shown that semantic categories (e.g. living versus non-living) play a role in the organization of our semantic memory [34, 126]. First, a broad clustering of related words into semantic categories can be obtained. Then, each category can be studied in more detail later. In this early study, we focused on obtaining such a broad clustering, employing the idea behind block modeling of graphs [178].

Words belonging to the same category are assumed to have strong relationships with each other (within-category word-pairs) and weak relationships with

words from different categories (between-category word-pairs). A within-category word-pair, when sequentially presented to a subject, should evoke a priming effect [1, 38]. This priming effect should in turn manifest in the amplitude of the N400 ERP component [6, 99, 102, 103]. Starting from a random assignment of words to groups, the solution can be optimized by moving words between groups in order to minimize the N400 amplitude in response to within-group word-pairs and maximize it in response to between-group word-pairs.

In the context this study, this means that the score given to a particular clustering depends on the N400 amplitude of all within-group word-pairs versus the N400 amplitude of all between-group word-pairs. When a word is moved from one group to another, the score given to the new clustering depends on all the word-pairs involving this word. This means that the decision to move a word between groups is never based on the N400 amplitude in response to a single word-pair.

We evaluated the method on a list of words that can be neatly clustered into two groups, namely animals versus furniture items.

4.3 Methods

4.3.1 Participants

The experiment was performed using 16 subjects (an additional two had to be discarded due poor sensor contact quality and one due to excessive eye blinks), of which 10 were male and 6 female, in the age range of 20 to 58 (mean 38, std. 11.06), all but one were right handed, 6 were native speakers of Belgian-French and the other 10 native speakers of Belgian-Dutch (the stimuli were presented in the subject's native language). This study was approved by the UZ Leuven ethics committee. All subjects were volunteers and signed an informed consent form before the experiment.

4.3.2 Materials

Word-pairs were formed by using all possible prime–target combinations (182) of the 14 words given in table 4.1. The list contains African animals and common furniture items. The stimuli were presented in the native language of the subject.

EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System (BioSemi, Amsterdam, the Netherlands), having a 5th order frequency filter with a pass band from 0.16 Hz to 100 Hz, and sampled at 2048 Hz. An EOG was recorded simultaneously and used to reduce eye artifacts in the EEG using the procedure outlined in [131]. Two electrodes were placed on both mastoids and their average was used as a reference for the EEG.

Dutch	French	English
bed	lit	bed
bureau	bureau	desk
deur	porte	door
giraf	girafe	giraffe
kast	placard	closet
leeuw	lion	lion
neushoorn	rhinoceros	rhinoceros
nijlpaard	hippopotame	hippopotamus
olifant	éléphant	elephant
stoel	chaise	chair
tafel	table	table
tijger	tigre	tiger
zebra	zèbre	zebra
zetel	canapé	couch

Table 4.1: Words used in the clustering experiment. The stimuli were all possible word-pair combinations of these words.

4.3.3 Experimental procedure

The stimulation paradigm and subject task were identical to the one used in chapter 2 and chapter 3, with the addition of proper counter-balancing of responding hand and assignment of the buttons to yes/no responses.

Subjects were seated in an upright position approximately one meter from a computer screen. The dominant hand rested upon the table with the index and middle fingers resting on mouse buttons. A trial consisted of the sequential presentation of a single word-pair. The first word of the word-pair (the *prime*) was presented for 200 ms and the second word (the *target*) for 1000 ms with a SOA of 500 ms. Finally, a response cue was given which prompted the subject to give a response by pressing a button.

Following the advise of Renoult et al. [105] regarding obtaining a semantic priming effect even when stimuli are repeated, the subjects were asked to determine whether the cue and target words belonged to the same semantic category. The subject responded by pressing one of two mouse buttons. The order of the mouse buttons and the hand used to operate the mouse was counterbalanced independently:

- 25% of the subjects operated the mouse with their right hand, using the index finger to indicate “yes” and the middle finger to indicate “no”.
- 25% of the subjects operated the mouse with their left hand, using the index finger to indicate “yes” and the middle finger to indicate “no”.
- 25% of the subjects operated the mouse with their right hand, using the middle finger to indicate “yes” and the index finger to indicate “no”.

- 25% of the subjects operated the mouse with their left hand, using the middle finger to indicate "yes" and the index finger to indicate "no".

Due to the nature of the stimuli, it is likely that the strategy of the subjects was to decide whether the two words belonged to the same category (animal versus furniture item).

4.3.4 Data preprocessing

The preprocessing procedure was identical to the one used in chapter 3. The EEG was bandpass filtered offline between 0.1 Hz and 50 Hz by a 4th order zero-phase IIR filter to attenuate large drifts and irrelevant high frequency noise, but retain eye movement artifacts. The EOG was used to attenuate eye artifacts from the EEG signal using the regression method outlined in [131]. After the EOG correction procedure, the signal was band pass filtered again, between 0.5 Hz and 15 Hz by a 4th order zero-phase IIR filter. Individual trials were obtained by cutting the continuous signal from 0.1 s before the onset of each target stimulus to 1.0 s after. Baseline correction was performed using the average voltage in the interval before the stimulus onset as baseline value. Before applying any multivariate analysis methods, the signal was further downsampled to 50 Hz to reduce the dimensionality.

To produce estimations of the N400 amplitude, the stLCMV beamformer, described in chapter 3, was used. This beamformer will produce estimations of a specific ERP component, given a template of the shape of this component. The template for the N400 that was constructed in that chapter (figure 3.6) was reused. This means the N400 template was derived from data that was recorded from different subjects than the ones used in this study.

The beamformer was used to produce estimation of the N400 amplitude for each trial. These estimations were averaged across all subjects to obtain a single measurement for each word-pair.

4.3.5 Block modeling based on N400 amplitude

Each of the 14 words was assigned to a group (an act referred to in this chapter as *clustering*). A score was assigned to each clustering of the words, derived from the N400 amplitudes. The idea is that words belonging to the same group produce small N400 amplitudes when presented as a word-pair and words belonging to different groups produce a large N400 amplitude. To reduce the search space, the evaluated clusterings were limited to two groups of equal size.

Let W denote the set of all 14 words, $A \subset W$ the set of words assigned to the first group and $B \subset W$ the set of words assigned to the second group, with the restriction that $A \cap B = \varnothing$ and $|A| = |B| = 7$. Let n_{ab} denote the estimated N400 amplitude for word-pair (a, b), where $a \in W$, $b \in W$, $a \neq b$. The scoring function was as follows:

$$f(A,B) = \sum_{a \in A} \sum_{b \in B} \begin{cases} -n_{ab} & \text{if } (a \in A \wedge b \in A) \vee (a \in B \wedge b \in B), \\ n_{ab} & \text{otherwise.} \end{cases} \tag{4.1}$$

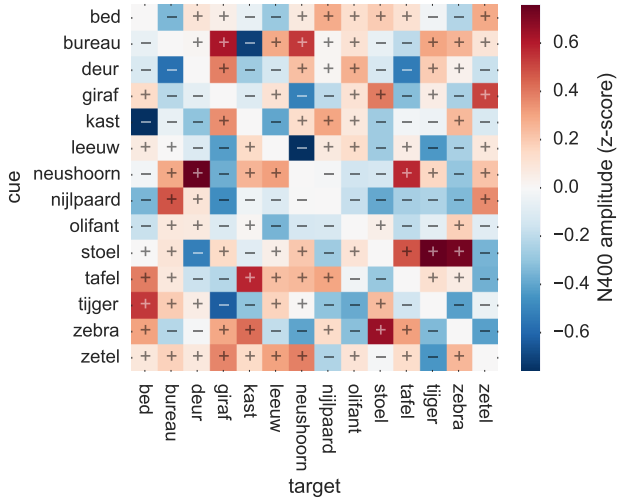


Figure 4.1: Estimated N400 amplitude for each word-pair. Words used as cue are given along the y-axis and words used a target along the x-axis. For each cue-target combination, the mean N400 amplitude, as estimated by an stLCMV beamformer, is shown as a colored square. N400 amplitudes are normalized to have a row-average of zero. Plus and minus signs indicate whether the N400 amplitude was higher or lower than the row average.

Hence, the clustering with the largest score is the one with the smallest *within group* N400 amplitude and the largest *between group* amplitude.

4.4 Results

The N400 amplitude estimates for each word-pair, averaged across subjects, are presented in figure 4.1 as a matrix, where each cell represents a cue-target combination.

Each possible way of clustering the words into two groups of equal size was evaluated. The highest score was achieved in the following case:

group 1: BED, DESK, DOOR, CLOSET, CHAIR, TABLE, COUCH

group 2: GIRAFFE, LION, RHINOCEROS, HIPPOPOTAMUS, ELEPHANT, TIGER, ZEBRA

In figure 4.2, The scores are presented in a matrix, with the order of the rows based on the optimal clustering as would be done in block-modeling. It can be noted that there are more high N400 amplitudes (red squares) between words that belong to different groups (off-diagonal quadrants) than between words that belong to the same group (diagonal quadrants). Also, there are more high N400 amplitudes in the furniture group (bottom-right quadrant) then there are in the animal group (top-left quadrant), which might indicate that the animal group is more coherent than the furniture group.

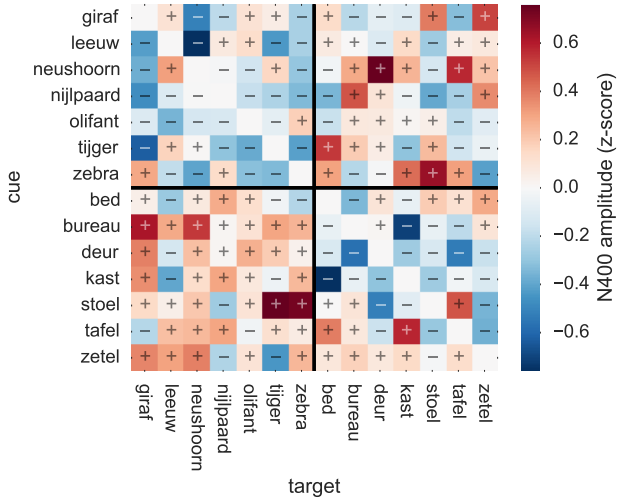


Figure 4.2: Optimal clustering by the block model. The rows and columns are ordered according clustering with the highest score. Black lines divide the word-pairs into quadrants, where the diagonal quadrants contain within-group pairs and the off-diagonal contain between-group pairs. Plus and minus signs indicate whether the N400 amplitude was higher or lower than the row average.

The scores of every possible clustering are presented in figure 4.3A. In order to have some insight into the behavior of the scores, they are laid out following the number of animals in the first group. For the general case of having k animals in the first group, there are $0.5\binom{7}{k}^2$ ways to perform the clustering. Note that since any animal not in the first group is necessarily in the second group, the cases $k = 0$ and $k = 7$ are equivalent, as are $k = 1$ and $k = 6$, etc.

There is the question of reliability of this result. If we were to repeat this experiment, how likely would it be the same clustering was found? Also, how many subjects are required in order to obtain a stable clustering? In order to answer such questions, we employed cross-validation. The scores for each clustering were evaluated on the mean N400 amplitudes of n subjects. Now, each clustering has $\binom{16}{n}$ scores, one for each possible way to select n out of the 16 subjects. Figure 4.3B (solid line) shows the percentage of times the animals versus furniture clustering was ranked the best. When no averaging is performed across subjects, this clustering was never obtained (likely due to too much variability in the N400 amplitudes). When averaging across 15 or more subjects, the correct clustering was always obtained. If we are willing to be lenient and also accept clusterings where some animal and furniture items were misplaced (black and gray dashed lines), the required number of subjects is lower.

The grand average ERPs, obtained by assigning the labels "within group" and "between group" based on the clustering with the maximum score, are presented in figure 4.4. To determine whether the amplitude of the N400 differs significantly between the classes, an LME model was used. The N400 amplitude,

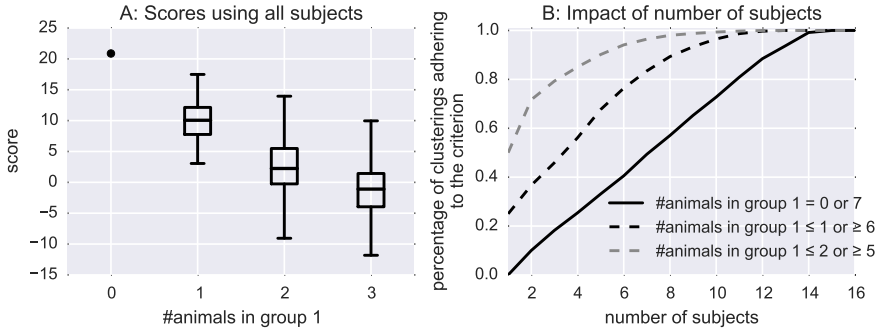


Figure 4.3: Scores for all possible ways to cluster the words into two groups of the same size. The scores are laid out on the x-axis in terms of the number of animals present in the first group. At the left and right boundaries of the axis are the two clusterings corresponding to all animals versus all furniture items. The scores for clusterings that mix animals and furniture items are represented by box plots, where the whiskers extend to the entire range of values. **A:** The scores of all the possible clusterings, computed using all subjects. **B:** The measurement was cross-validated using only n subjects. This graph shows, out of all possible combinations to select n subjects, the percentage of times a correct clustering was obtained (solid line), two words or less in the wrong cluster (black dashed line) or four words or less in the wrong cluster (gray dashed line).

as measured by the stLCMV beamformer, was entered as independent variable, with as dependent variable a $\{0, 1\}$ dummy encoding of the class labels. Subjects (random slopes only) and target words (random intercepts only) were entered as random effects. The model found a significant difference of N400 amplitude between the classes ($t(2897) = 7.129,\ p < 0.0001$). In addition to the N400, a P600 can be observed on the frontal sensors, which is commonly observed when stimuli are repeated [179].

4.5 Discussion and conclusion

This experiment represents a step forwards in our ability to map the semantic relationships between words, as utilized by our semantic memory.

In traditional priming experiments involving category membership, word-pairs are presented to a subject that are known a priori to belong to the same semantic category. The research question is in this case whether the N400 amplitude is affected by category membership, which has been found to indeed be the case. Our experiment turned this paradigm upside down. Starting from the now known fact that N400 amplitude is modulated by category membership, our research question was which stimuli belonged to the same semantic category.

It is important to note that N400 amplitude is affected by many properties of a stimulus (figure 3.5). In this study, that uses cue–target pairs, we are interested in the N400 response to the target word. This is influenced both by properties of the target word and, importantly, properties of the relation between the cue

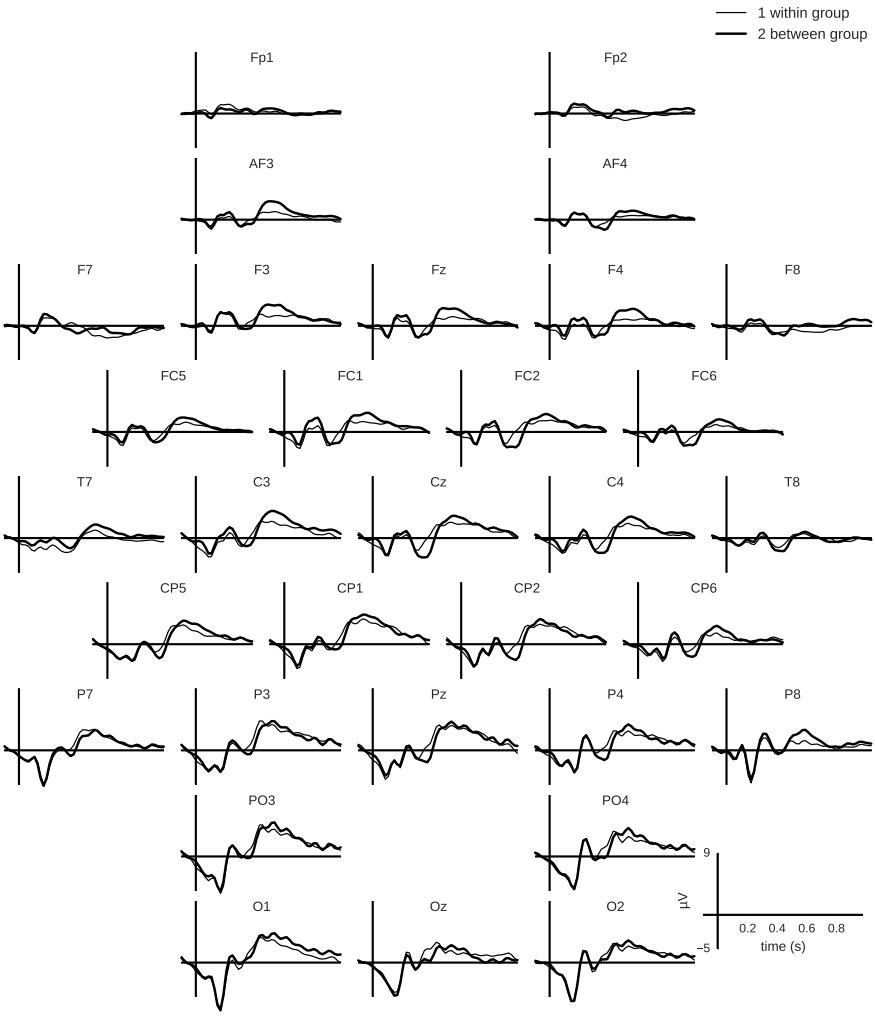


Figure 4.4: Grand average ERPs. The labels "within group" and "between group" labels are based on the clustering of the words with the highest score.

and target (e.g. whether the words belong to the same semantic category). In the clustering approach described in section 4.3.5, the decision to assign a word to a group is based on how the N400 amplitude changes when the target word is combined with different cues. If we assume that the effect of target-specific properties on the N400 amplitude, such as frequency or age-of-acquisition, is not affected by the prime word, these effects are properly counterbalanced.

The fact that the semantic category (animals versus furniture) of the stimuli could be reconstructed, based on N400 amplitude estimations alone, opens the door to start evaluating more complex relationships between words. For example, a follow-up study could be to study how animals are clustered within our semantic memory. Possible ways to cluster them include taxonomic categories (mammals, fish, ...), predator–prey, foreign–local fauna, etc. Instead of designing separate experiments for each possibility, our method can indicate the type of relationship that results in the strongest semantic priming effect, given an unordered list of animal names.

Chapter 5

Discussion and conclusions

5.1 Summary of contributions

Several methods have been introduced to obtain a clear recording of the N400 potential and produce an accurate estimate of its amplitude. These methods remove some of the obstacles the N400 to be a useful measure of relationship strength between words. We closed by taking a first step towards using the N400 to map out semantic relationships, as they are utilized by our semantic memory.

In chapter 2, we discussed some pitfalls when designing experiments, aimed to capture a specific brain process. As generally many ERP components are evoked by a rich stimulus such as a written word [58], care must be taken in the experimental design to make sure the amplitude estimation of the COI is not influenced by fluctuations of these interfering components. An ideal experimental design avoids actions that produce EEG signals that are unrelated to the brain process of interest.

In many semantic priming experiments, it is of interest to capture both RT measurements as well as the ERP, since priming can be shown through both measurements. For example, figure 2.3B and table 2.1 demonstrate the relationship between RT and FAS, while figure 2.3C and figure 3.4 demonstrate the relationship between N400 and FAS. However, to obtain meaningful RTs, the subject has to perform a speeded task. We have shown that during a speeded task, the actions of selecting a button to press and the actual pressing of the button produces P3s and MRPs that overlap with the N400. Most problematic was the latency of the P3b, which closely follows the onset of the button press and is therefore correlates with the N400 amplitude. The severity of this interference was demonstrated in section 2.4.5, where a difference in P3b-latency could completely mask the differences in N400 amplitude.

Where others merely mention the problems of a speeded response task in passing [14, 74, 76], we have demonstrated the severity of the problem. Modifying the experimental paradigm to postpone the button press of the subject by 1 s proved effective in removing the overlap of said ERP components.

Traditionally, the voltage of the ERP waveform, averaged across some ROI, is taken as estimation of component amplitude [55], which was used in chapter 2 as well. Even with the elimination of the P3 and MRPs, components still remain that may confound estimations of N400 amplitude if such a naive method is employed.

In chapter 3 we explored the use of linear spatio-temporal filters to isolate a COI from the rest of the signal. These filters adapt to the data with a certain autonomy: instead of specifying the filter weights directly, the researcher specifies

the parameters of a search space, in which the filter weights represent the solution that satisfies some criterion. This autonomy, if left unchecked, may lead to a "black box" approach where the researcher looses track of what the output of the filter ultimately represents. In order to come to a pragmatic approach to design a filter to isolate the N400 component (or any specific COI), we introduced two performance criteria:

1. the output correlates well with the actual amplitude of the COI (sensitivity)

2. the output does *not* correlate with any structured interfering signals, such as other ERP components (specificity)

In light of these criteria, we propose to use an LCMV beamformer, slightly extended from its original formulation as spatial filter to act as spatio-temporal filter (stLCMV). This brings the beamformer out of its traditional context of source reconstruction of EEG and MEG signals into a new context as flexible filter to isolate ERP components. In a study on both simulated and real EEG data, we evaluated the stLCMV beamformer along with the traditional ROI-mean and some popular supervised learning approaches (figure 3.3 and figure 3.5). We discussed the merits and weaknesses of the approaches in terms of the two performance criteria. In terms of accuracy of its estimation of COI amplitude, the first criterion, the stLCMV beamformer's performance is comparable with that of well known supervised learning techniques, such as the lSVM. Where the stLCMV beamformer shines is its remarkable robustness to interfering ERP components, the second criterion, which is a favorable quality if its output is used as an estimation for the amplitude of a COI. Variance in the output of the beamformer that cannot be explained by N400 amplitude, was more likely to be due unstructured noise, which is a desirable property during statistical analysis. It allows the researcher to have confidence that the filter's output is an actual measure of the COI amplitude and not something else.

In chapter 4, the stLCMV beamformer was put to work to produce N400 amplitude estimates in a dataset, where the N400 effect is quite small. Due to the repetition of stimuli, not only was the N400 amplitude small, an interfering P600 component was also produced. In order to produce a clear template of the N400 without any interference from the P600, we re-used data from a previous recording where stimuli were not repeated. This highlights the beamformer's attractive property of explicitly defining the signal that will be isolated. The template of this signal does not have to be inferred from data of the current recording, but can be constructed using any method deemed appropriate by the researcher.

The main objective laid out in section 1.5 was to devise a measure of relationship strength between words, based on the amplitude of the N400 of a subject in response to the words in a semantic priming paradigm. With the combination of the experimental paradigm described in chapter 2 and the stLCMV spatio-temporal filter described in chapter 3, we believe to have laid the groundwork for an adequate measure. This measure was shown in chapter 4 to be able to reliably group words according to category membership.

5.2 Experimental paradigm and the P3

In this thesis, some new tools are presented to study semantic priming. During the development of these tools, it was preferable to work with data which show a clear priming effect. To this end, written prime–target word-pairs were chosen as the stimulus modality. They are easy to construct, can be presented very fast (a typical trial lasted around 3 s, enabling the collection of many trials during a single recording) and allow for good control over the onset of its presentation (the stimulus itself has no time dimension, like a spoken word or animation). The stimuli were presented using a large SOA of 500 ms to allow full semantic processing of both the prime and target stimuli, which produces a larger priming effect than sorter intervals.

The stimuli were combined with a JAM task, which has the advantage of making the prime relevant to the task and promote full semantic processing of both the prime and target stimuli. Renoult et al. (2011) [105] show that this is important when stimuli are repeated, which is the case in chapter 4. In their study, they show that a frequently used alternative task, lexical decision, fails to produce a reliable N400 effect when stimuli are massively repeated. Chapter 4 reproduces one of their results, namely that with a JAM task, the N400 effect can be observed, even when stimuli are repeated.

Unfortunately, Renoult's study employed a speeded button response task. In chapter 2, we demonstrated that a speeded design evokes a P3 component, which latency correlates with the RT of the subject. Indeed, in the results of Renoult, such a component can be observed. With figure 2.3, we demonstrate how easy it is to confuse the "walking" P3 with the stationary N400. Therefore, the question is how much of Renoult's effect is caused by changes in P3 latency and how much by N400 amplitude fluctuations.

In chapter 4, the subject's response was delayed and based on the conclusions drawn in chapter 2, we would suspect that the P3 was suppressed. However, the failure to detect the P3 in the delayed condition in thay study (figure 2.5) may be due to limitations in the detection method. In chapter 4, a P3 component can still be observed (figure 4.4). Even when the actual button-press is delayed, the decision about the relatedness of the words is still made during the time window of the N400 and may produce a P3.

Perhaps it is possible to design a task that would further reduce the interference of the P3 component. Tom Heyman et al. (2013) [18] propose to hide one of the letters of the target word. Some time after the presentation of the target word, a letter is displayed for which the subject has to decide whether it is the missing letter. Like the JAM task, the prime stimulus is relevant for the task (a related prime helps the subject guess the target word). But unlike the JAM task, the letter decision task cannot be performed until the missing letter is displayed, which is well after any N400 effects. As Heyman's study employed RT measurements only, an ERP study would have to be performed to explore the P3 component generated during this task.

However, semantic priming is a phenomenon that is characterized by an increased speed in the processing of a stimulus. It may be the case that, regardless

of the experimental paradigm, a "walking" P3 component will always be generated. This stresses the importance of using proper multivariate techniques if one wants to study the N400 and P3 in isolation. The ability of a filter to do this is expressed by the second performance criterion in chapter 3. A criterion on which the stLCMV beamformer performs well.

5.3 Simulation study comparing multivariate filtering approaches

In the supplementary information of chapter 3, a multitude of approaches to multivariate filtering are evaluated using simulated data. In the data simulation, four parameters were tweaked that are designed to demonstrate the strengths of the approaches and expose their weaknesses.

In chapter 3, the task for each filtering approach was to isolate COI from the rest of the signal. The filters could use training examples of trials with and without the COI, simulated to originate from 10 different subjects. The performance of the filters was evaluated by splitting the variance of their output into three parts:

1. variance attributable to the true COI amplitude
2. variance attributable to amplitude modulations of other ERP components that are not the COI
3. variance attributable to non event-related EEG activity and sensor noise

Ideally, we would like the output to correlate perfectly with COI amplitude. For example, when designing a filter to isolate the amplitude of the N400 component, the optimal filter would give a completely accurate measurement of this. However, any variance not attributable to the COI, we would like to be uncorrelated with the amplitude of other ERP components. For example, in chapter 2 we found that there is a P3 component that is overlapping with the N400. We would like the output of the N400 filter to be uncorrelated with this P3 component. Any variance in the output of the filter not attributable to the COI or other ERP components, must in the simulation necessarily be correlated with the ongoing EEG and sensor noise.

5.3.1 Unstructured noise

The first parameter that was tweaked was the amount of unstructured noise in the signal. This noise emulates ongoing, non event-related, brain processes. It was simulated by a set of dipoles which randomly popped in and out of existence. This noise can be characterized as spherical: there is no linear projection orthogonal to this noise and therefore linear filters cannot counter it.

The lSVM, LR and shrinkage-LDA algorithms are full spatio-temporal filters, which means they have to learn $m \times n$ (#electrodes $\times$ #samples) parameters. To combat overfitting, LR and shrinkage-LDA employ $\ell 2$ regularization directly, whereas the lSVM's c-parameter (the region surrounding the decision boundary in which training examples are ignored) works differently, but can serve the

same purpose. With low levels of unstructured noise, these algorithms perform well, however, when the amount of unstructured noise is increased, they start to overfit.

Three approaches are explored to reduce overfitting:

First, by reducing the number of features. The restricted-lSVM performs the analysis on several predefined electrodes and time points (the ROI), which are known to optimally capture the COI. The algorithm starts to perform better than the unrestricted version as unstructured noise increases.

Second, by using more training data. The group-lSVM not only used the training data of the subject, but from all subjects. As long as the records of the subject are similar, this will perform better than the version that only trains on the data collected from the current subject.

Third, by reducing model complexity. By combining separate spatial and temporal filters, which learn m and n parameters separately, the original $m \times n$ parameter space is reduced to $m + n$. The Fisher criterion (FC)+lSVM approach first designs a spatial filter, based on the Fisher criterion [117] and filters its output with an lSVM. However, ERP components seem to be difficult to capture in such a scheme and the algorithm only achieves a minuscule performance gain over the spatio-temporal lSVM when unstructured noise increases. The bCSP combines two CSP filters (one spatially and one temporally) to achieve the same effect. However, this method is outperformed by all the other filters.

The top scores in this case to go the beamformer techniques. The amount of overfitting displayed by the stLCMV beamformer approach depends on manner in which the template of the COI is constructed and the quality of the covariance estimation. The perfect-stLCMV and perfect-chained-LCMV beamformer cheat by being supplied with the true template for the COI. The latter algorithm, which applies separate spatial and temporal filters, outperforms the first, due to the lower number of elements in the covariance matrix that need to be estimated.

The versions of the beamformer that do not cheat (stLCMV and chained-LCMV) have equal performance with their cheating counterparts. This means that the method used to construct the template is very robust against unstructured noise. By enforcing zeros in the template for time samples outside the predefined ROI, the influence of random, meaningless fluctuations is reduced. Since the template is constructed on the grand-average of all the training data, the beamformer effectively makes use of all the training data, just like the group-lSVM. The template for the ss-stLCMV beamformer is estimated by using only the training data for the current subject. It is outperformed by the version that uses all training data.

5.3.2 Structured noise

This noise emulates ERP components that possible overlap with the signal of interest. These were simulated by a fixed set of dipoles with fixed locations in the head model and a fixed temporal activation pattern. The only thing randomized from trial to trial was the scaling of the temporal activation pattern (y, see figure 1.5). This causes them to generate signals that behave as stable ERP com-

ponents, with a random amplitude for each trial. Each simulated recording had a unique set of structured noise dipoles. This means that analyzing the structured noise in one recording would not give information about the structured noise in another. Given enough training data for the current recording, linear filters should be able to find a projection that is orthogonal to this type of noise and effectively counter it.

By tweaking the amount of structured noise, the algorithms are tested on their ability to isolate the COI from other ERP components. This means that the variance that is explained by structured noise is an important metric. The clear winner on this metric is the stLCMV beamformer. Its output never correlates strongly with other ERP components, even when the EEG signal is completely dominated by them. This illustrates the main reason why a researcher would want to employ the beamformer instead of another algorithm such as an lSVM.

When the structured noise reaches high levels, care must be taken when designing the template for the beamformer. Where the output of the cheating perfect-stLCMV beamformer is not influenced by the presence of structured noise at all, the accuracy of the stLCMV beamformer suffers slightly. Especially the ss-stLCMV approach, which doesn't take all training data into account, fails to estimate a good template of the COI, in the presence of large amounts of structured noise. However, the decrease in performance presents itself as an increase in correlation with the unstructured noise. The amount of correlation with other ERP components remains low, which is a desirable characteristic.

The chained-LCMV beamformer performs poorly, along with the other approaches that apply separate spatial and temporal filters (FC+SVM and bCSP). Where their low complexity was an advantage in the face of unstructured noise, their inability to model electrode-time interactions is now a disadvantage.

The lSVM, LR and shrinkage-LDA algorithms all perform well in the presence of overlapping ERP components. Their output is generally not affected by the amount of structured noise. However, their output correlates more with other ERP components than the stLCMV beamformer.

The restricted-lSVM now has a slight disadvantage over the unrestricted variant. Restricting the number of features hinders its ability to counteract structured noise, leading to a somewhat lower performance. The performance difference is not much though and given its superior performance in the presence of unstructured noise, the restricted-lSVM might still be preferable.

Since the structured noise is unique to each subject, the group-lSVM is at an disadvantage. Unlike the beamformer approaches, which only assume that the COI is similar across subjects, the group-lSVM assumes that both the COI and all other ERP components are similar. This assumption is violated in the simulation to demonstrate this weakness of the group-lSVM and demonstrate the difference with the beamformers, which also pool together training data, but are still capable of adapting to the unique noise sources in a recording.

5.3.3 Temporal jitter

The COI was simulated as a single, fixed, dipole, which peaked on average at 400 ms. To simulate small changes in the COI between subjects, some randomness was applies to its latency. The temporal jitter parameter controlled how much earlier or later the COI was allowed to peak, relative to 400 ms. The challenge for the filtering approaches is to capture the uniqueness of the COI in each subject.

The filtering approaches that use only the training data of the current subject are at an advantage here. Their performance is not affected by changes in the COI. This is also the case for the restricted-lSVM and ss-stLCMV approaches, as long as the COI still falls in the ROI. Since the COI is now no longer similar for all subjects, pooling training data is no longer a good strategy. The group-lSVM is now outperformed by the subject specific lSVM and restricted-lSVM variants.

The beamformer filters can be thought of as a spatio-temporal notch filter. They pass only signals that adhere to the template, and preferably nothing else. In the case of the perfect-stLCMV and perfect-chained-LCMV beamformers, the template is fixed at 400 ms and not inferred from the training data at all. Since due to the temporal jitter, the COI no longer fits this template, it could now be regarded as a noise ERP component that happens to be in the exact spatial location as the COI with only a minor difference in latency. As such, it can demonstrate the sensitivity of the beamformer filters (the width of the "notch" if regarded as a notch filter). It can be seen that ERP components, even with peak-latency of 200 ms away from the COI, still influence the output of the filter if they happen to be in the exact same spatial location as the COI. To effectively remove ERP components, both their spatial location as well as their latency in time should be distinct from the COI.

5.3.4 Label accuracy

With simulated data, the training labels can be perfect. However, in real EEG data, training labels come with some level of uncertainty, as we generally don't know a priori the exact amplitude of the COI for each trial. This parameter tweaks the reliability of the training labels by making them the weighted average of the true COI amplitude and the amplitude of some other ERP component. This way, we model situations where the researcher aims with his/her experimental manipulation modulate only one components, but the manipulations actually models multiple (e.g. it modulates both N400 and P3 amplitude). In the context of this thesis, we would like the algorithm to isolate only the COI, even though we know our experimental manipulation is likely to affect multiple components.

This parameter was included to demonstrate the main weakness of the lSVM, LR and shrinkage-LDA approaches. When the training data is not perfect, the performance of these algorithms drops rapidly. Understandably, they isolate any ERP components that help to distinguish between classes, irregardless of whether they are the COI. This is illustrated by comparing the variance of the output explained by the COI versus the output explained by other ERP compo-

nents. As the accuracy of the training labels stop correlating the with COI and start to correlate with another ERP component, the output of the algorithms follow suit.

Restricting the features to a ROI around the COI helps, so the restricted-lSVM approach outperforms the unrestricted variant. This depends on whether the training labels follow an ERP component that falls within the ROI or not (hence the large standard error margin for this algorithm).

As always, the performance of the beamformer techniques depends on the accuracy of the template. The cheating versions with perfect templates are not dependent on the training data at all and their performance is not influenced by this parameter. The correlation between the output of the non-cheating beamformers and the COI decreases when the training labels can no longer be used to create an accurate template. However, the estimation of this template is quite robust against inaccuracies of the training labels, making the stLCMV and chained-LCMV beamformers beat the other filtering approaches on this metric. Most importantly, the variance of the output of the beamformer filters that can be attributed to ERP components that are not the COI remains low, even when the training labels correlate with a competing ERP component. This means that even though we know our experimental manipulation is likely to affect multiple components, the beamformer can be used to isolate a specific component.

5.4 EEG study comparing multivariate filtering approaches

The evaluation of the various multivariate filtering approaches on real EEG data has an important limitation: the true amplitude of the COI (i.e.the N400) is unknown. Instead, the evaluation necessarily has to be indirect through the usage of known correlates with the N400 (table 3.2). Two metrics were used in this evaluation.

The first metric is the magnitude of the correlation between the estimated N400 amplitude and stimulus properties that are known to correlate with the N400. In the case of the FAS property, this measures the ability of the filtering approach to reconstruct the training labels. Because the stimulus properties used in this study are inter-correlated (figure 3.2) a high correlation with FAS also means a high correlation for other properties. The chained-LCMV beamformer scored best on this metric, marginally beating the lSVM (figure 3.5). However, during the simulation study it was found that the output of both the chained-LCMV and the lSVM approaches is possibly influenced by other ERP components as well.

The second metric is the total number of significant correlations found. This tests whether the filter generalizes beyond reproducing FAS scores from the ERP. The stLCMV beamformer scored best on this metric. This, combined with the simulation results, lead to the conclusion that although other filters might produce stronger correlations with FAS, the stLCMV beamformer more accurately isolates the true N400 component.

While the combination of the simulation study and study on real EEG data are sufficient to prove the points made in chapter 3, a shortcoming of the latter

is that no correlates with other components (P150, N200, P3, see section 1.3.1) were included in the study. If a follow-up experiment were to be conducted, it would be useful to include stimulus properties that correlate with these components. The researcher could then verify that the filter output does not correlate with these properties.

5.5 Practical considerations concerning the LCMV beamformer approach

The effectiveness of the beamformer filter is dependent on two factors: the accuracy of the template and the quality of the estimation of the covariance estimation.

5.5.1 Designing a proper template

One of the key factors influencing the behavior of the beamformer filter is of the template given as input. There are many ways in which this template could be designed.

In section 3.3.5, the spatio-temporal template for the N400 was designed by taking the outer product of a separate spatial and temporal template. The advantage of this approach is that the spatial template can be used to design a spatial beamformer, which in turn can be used to filter a large part of the noise and construct a relatively clean temporal template. The disadvantage of this approach is that it is unable to account for changes in the spatial distribution of the component over time. For some components, especially ones which spatial distribution changes over time, better results may be obtained by creating an effective contrast and simply using the difference wave as template.

Alternatively, one can employ multivariate techniques (either supervised or unsupervised) that simultaneously create patterns in addition to filters (e.g. ICA [109], CSP [180] and xDAWN [116]). The patterns produced by such techniques could in turn serve as basis for the template used by the beamformer.

Yet another alternative approach would be to avoid using EEG data for creating the template and instead rely on other modalities, such as MEG and fMRI. The template of the COI could be constructed through a realistic head model, like the one used in section 3.3.6. To accurately place the dipole (or dipoles) that generate the COI, information could be used from source localization studies [91, 92, 94].

5.5.2 Estimating the covariance matrix

Whereas the template models the COI, the beamformer uses an estimation of the covariance matrix of the EEG signal to model any noise sources. Hence, the performance of the filter also depends on the accuracy of this estimation.

Robustly estimating a covariance matrix has been a thoroughly studied topic [158–160]. Although the basic formula for obtaining an (unbiased empirical) es-

timate of the covariance matrix of a given matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is straightforward:

$$\Sigma_X = \frac{1}{n-1}\mathbf{X} \cdot \mathbf{X}^\intercal \tag{5.1}$$

this estimate becomes unreliable when either the number of variables (m) is larger than the number of observations (n) or when there is a large amount of unstructured noise.

A commonly used tactic to increase the reliability of the estimate is to employ shrinkage towards an identity matrix:

$$\Sigma_X = \lambda \mathbf{I} + (1-\lambda)\left[\frac{1}{n-1}\mathbf{X} \cdot \mathbf{X}^\intercal\right], \tag{5.2}$$

where $\mathbf{I} \in \mathbb{R}^{m \times m}$ is an identity matrix and λ defines the amount of shrinkage to apply.

The shrinkage parameter encodes our belief of how much equation 5.1 underestimates the amount of unstructured noise in the signal. For example, setting $\lambda = 1$ implies that the signal wholly consists of white noise and no correlation exists between the recorded EEG samples. In this case, the LCMV beamformer filter is reduced to a matching filter [107], where the filter weights are equal to the supplied template (scaled by a constant):

$$\mathbf{w} = \frac{\Sigma_X^{-1}\mathbf{a}}{\mathbf{a}^\intercal \Sigma_X^{-1}\mathbf{a}}, \quad \text{original LCMV formula} \tag{5.3}$$

$$\mathbf{w} = \frac{\mathbf{I}^{-1}\mathbf{a}}{\mathbf{a}^\intercal \mathbf{I}^{-1}\mathbf{a}}, \quad \text{substituting } \Sigma_X = \mathbf{I} \tag{5.4}$$

$$\mathbf{w} = \frac{\mathbf{a}}{\operatorname{var}\mathbf{a}}. \tag{5.5}$$

Decreasing λ allows the beamformer filter to capture more structured noise and more aggressively correct for things such as interfering ERPs and EOG artifacts.

To estimate the optimal shrinkage parameter, one can employ a cross-validation scheme and check the performance of the resulting filter against the training labels. However, this requires accurate training labels, which in the case of the N400 are difficult to obtain. Therefore, we have chosen to estimate λ by using the OAS algorithm [158] (an improved version of the Ledoit-Wolf estimator [181]) that has been shown to perform well in the case of beamforming [158].

5.6 Clustering words

In chapter 4, words were clustered together based on a metric that was derived from the N400 amplitude, using the stLCMV beamformer.

Since this study represents a first step in uncharted territory, an important limitation was put into place to reduce complexity. The search space for the optimal clustering was restricted to solutions that consist of two clusters of equal size. This restriction enabled an exhaustive search of the entire solution space,

sidestepping the complicated optimization strategies required when block models become more complex (e.g. [182]). Another convenient consequence of the enforcement of equal group sizes is that the resulting within-category versus between-category contrast is always properly counterbalanced. If clusters of unequal sizes are allowed, the number of trials in the within-category group will be lower than the between-category group. Additionally, if the current scoring metric (equation 4.1) is not modified, words that intrinsically produce a larger N400 amplitude (e.g. because they have a low frequency, large age of acquisition, large orthographic neighbourhood, etc.) will be biased towards the smaller cluster.

5.7 The road ahead

We have developed a suitable metric for N400 amplitude during semantic priming and demonstrated a use case how it could be applied.

The logical next step would be to distill this into a metric for relationship strength between words as they are represented in the brain. This would involve taking into account the different factors that influence N400 amplitude (see section 1.3), either by careful counterbalancing of stimuli or by partial regression to compensate for confounding factors.

The new beamformer technique is a promising analysis tool for ERP research. There are many possible studies where averaging across many trials is not feasible, either because not enough trials can be collected or because the hypothesis is about the stimuli themselves. But even in traditional studies that do employ averaging trials, the beamformer can help improve accuracy and may reveal more subtle effects.

In section 3.3.5, we present a straightforward method for constructing a template of the COI, used by the beamformer. This approach can be improved in various ways. For example, one may choose to adapt the template slightly to the data of the current subject. For example, N400 latency is known to vary slightly from subject to subject [103]. This variation can be captured by sliding the template in time to optimize correlation with the subject's ERP. The tradeoff here is the more the template is allowed to adapt to the data of a specific recording, the more the beamformer method mimics supervised learning approaches.

In chapter 4, words were clustered together based on a metric that was derived from the N400 amplitude. The underlying idea is that the clustering represents some structure in the organization of concepts in our semantic memory. The two categories used in the study (animals versus furniture items) represents a rather extreme case. This was intentional, so that the result could be easily interpreted. It will be interesting to apply the method to dig out some more subtle relationships, where the optimal clustering is not known in advance. For example, a follow-up study could involve a set of animals, for which there are many ways to cluster them (predator–prey, mammal–fish, indigenous–exotic, etc.) and see which clustering "wins". Also, the block model employed in this study was the most simple possible. Inspiration for more details studies, employing more groups and also modeling the relationships between groups, can

be found in the literature on block modeling.

In the broad scheme of things, this thesis tried to flip the traditional method of conducting ERP studies upside-down. Traditionally, the researcher would come up with an hypothesis about the organization of our semantic memory or some other aspect of our semantic system, design stimuli to test this theory and then use ERP analysis to either confirm the theory (and publish it) or not (and usually not publish it). Chapter 4 starts from the ERP recordings and tries to work towards an hypothesis about the organization of our semantic memory from there.

The result might be the expected result (e.g. animals might not end up being clustered into all predators versus all prey) and sometimes not (e.g. the predators might end up being clustered with the prey animal they eat the most).

Summary

Semantic relationships between words are a key component in the functioning of our semantic memory, our ability to understand and produce language and our ability to reason about the world. For example, *dogs* have *tails*, which they *wag*, but only when they are *happy*. Constructing a map of these semantic relationships and their relationship strength (for example in the form of a semantic network) is useful in our quest to understand the aforementioned processes.

There are many ways to construct these maps, for example counting the number of co-occurrences of words in a text, using dictionaries and thesauri or simply asking people to write down word associations. However, if we want to study how the brain processes language, the more useful measures might be those based brain responses themselves. An experimental paradigm that is known to reveal automatic brain responses to semantic relationships is semantic priming. This semantic priming effect occurs when a target stimulus is preceded by one or more priming stimuli that share a semantic relationship with the target. In this case, the priming stimuli create a context that, through various mechanisms, facilitates the processing of the target stimulus by the brain.

To record brain activity, we employed electroencephalography (EEG), which is a non-invasive method that is widely used in the neurosciences to study ongoing processes in the living brain. One of its disadvantages is that the conductive properties of tissue between the origin of the signal and the electrodes on the scalp cause each electrode to pick up signals from a large region of the brain. If one is interested in studying one particular brain process, as is usually the case, these signals must be separated in order to isolate the process of interest. A commonly used technique to do so is the construction and analysis of the event-related potential (ERP). During semantic priming experiments, stimuli preceded by a related prime generate a smaller amplitude of the N400 component in the ERP then onces that are preceded by an unrelated one. Because the subject has no voluntary control over his/her N400 amplitude, this component could serve as a proxy measurement of the relationship strength between the representations of the stimuli in the brain.

The major challenge addressed in this thesis is to obtain an accurate measurement of the amplitude of the N400. Averaging EEG responses to hundreds of stimuli and constructing the ERP waveform is less of a problem when one wants to study the properties of the N400 itself. The focus of this thesis lies however on deducing the properties of the stimuli, namely the relationship strength between them. Therefore, averaging across stimuli becomes problematic and other solutions must be found to isolate the N400 from the rest of the signal.

A good way to prevent occlusion of the N400 by other brain processes is to design the experiment in such a way that other strong ERP components do not occur in the time window of interest. A common experimental design for studying semantic priming instructs the subject to perform a speeded decision

task and press the correct button. This has the advantage of capturing the response time of the subject as an extra measurement of the semantic priming effect. However, the decision making process and the pressing of the button evoke ERP components that unacceptably obscure the N400. We found that delaying the button response eliminates this overlap.

To further isolate the N400 from the rest of the ongoing EEG signals, a new multivariate approach called a spatio-temporal LCMV beamformer, was introduced. Using both simulated and real EEG signals, the performance of the beamformer was evaluated, next to traditional mean voltage measurement and supervised learning approaches, such as the lSVM. We found that the performance of the beamformer and lSVM were comparable in terms of correlation with actual N400 amplitude, but that the beamformer was more robust against influences from nearby ERP components. Variance in the output of the beamformer that cannot be explained by N400 amplitude, was more likely to be due unstructured noise, which is a desirable property during statistical analysis. It allows the researcher to have confidence that the filter's output is an actual measure of N400 amplitude and not some other component.

The combination of experimental design and the stLCMV beamformer filter resulted in a good measurement of N400 amplitude. This measurement was employed to perform a clustering study in which a list of words were clustered into groups. The groups were based on the notion that words belonging to the same group would produce a small N400 amplitude, when presented as a word-pair, than words belonging to different groups. Starting from an unordered list of 14, we showed that the N400 component could be used to reliably cluster them into their original categories: 7 African animals and 7 furniture items.

The work presented in this thesis clears to way to employ the N400 as a basis for measuring the strength of semantic word relations, as represented in the brain.

Samenvatting

Semantische relaties tussen woorden zijn een belangrijk component in het functioneren van ons semantische geheugen, onze capaciteit om taal te begrijpen en te produceren en onze capaciteit om te redeneren over de wereld om ons heen. Bijvoorbeeld, *honden* hebben een *staart*, waar ze mee *kwispelen*, maar alleen wanneer ze *blij* zijn. Het in kaart brengen van deze semantische relaties en hun sterkte (bijvoorbeeld in de vorm van een semantisch network) is nuttig in onze zoektocht naar het begrijpen van bovengenoemde processen.

Er zijn vele manieren om deze kaarten te maken, bijvoorbeeld door het tellen hoe vaak woorden samen voorkomen in een tekst, gebruik maken van woordenboeken en thesauri, of simpelweg door mensen te vragen woordassociaties op te schrijven. Wanneer we echter willen bestuderen hoe het brein taal verwerkt, is het wellicht nuttig om tot een maat te komen die afgeleid is van de hersenactiviteit zelf. Semantische priming is een experimentele opstelling waarvan bekend is dat het automatische hersenresponsen oproept die samenhangen met de semantische relaties tussen woorden. Dit priming effect vindt plaats wanneer een *target* stimulus voorafgegaan wordt door één of meerdere *priming* stimuli die een semantische relatie hebben met elkaar. In dit geval creëeren de priming stimuli een context die, door verschillende mechanismen, het verwerken van de target stimulus vergemakkelijken.

Om hersenactiviteit te meten maakten we gebruik van elektroencefalografie (EEG), wat een non-invasive methode is om de processen in een levend brein te bestuderen. Een van de nadelen van deze techniek is dat de elektrische geleiding in het weefsel tussen de bron van het signaal en de elektroden op de scalp ervoor zorgt dat iedere elektrode signalen opvangt van bijna alle delen van het brein. Wanneer men geïnteresseerd is in het besturederen van een specifiek proces in het brein, zoals vaak het geval is, dan moeten deze signalen geïsoleerd worden van de rest. Een veel gebruikte techniek hiervoor is het analyseren van de event-gerelateerde potentiaal (ERP). In semantische priming experimenten roepen stimuli die voorafgegaan worden door een gerelateerde prime een kleinere N400 component van de ERP op dan degenen die voorafgegaan werden door een ongerelateerde. Omdat de proefpersoon geen vrijwillige controle heeft over zijn/haar N400 amplitude, zou deze component bruikbaar kunnen zijn als een proxy meting voor de sterkte van de relatie tussen de representaties van de stimuli in het brein.

Deze thesis gaat over methoden om een accurate meting van de amplitude van de N400 te verkrijgen. Het middelen van EEG responsen van honderden stimuli en het opstellen van een ERP waveform is wellicht geen probleem wanneer de eigenschappen van de N400 zelf worden bestudeerd. De focus in deze thesis ligt echter op het achterhalen van de eigenschappen van de stimuli, namelijk de sterkte van de semantische relaties, dus is het problematisch om over stimuli te moeten middelen en zal er een andere methode moeten worden gevonden om

de N400 te isoleren van de rest van het signaal.

Een goede manier om de verstoring van de N400 door andere hersenprocessen tegen te gaan is om het experiment dusdanig te ontwerpen dat er geen andere sterke ERP componenten plaatsvinden in de desbetreffende tijdspanne. Een veelgebruikte experimentele opstelling voor het bestuderen van semantische processen is om de proefpersonen de instructie te geven om zo snel mogelijk een beslissingstaak uit te voeren en op de juiste knop te drukken. Dit heeft het voordeel dat de reactietijd van de proefpersoon ook wordt opgenomen, wat kan dienen als extra meting van het semantische priming effect. Het beslissingsproces en het drukken op de knop roepen echter hun eigen ERP componenten op die op onacceptabele wijze storen met de N400. We ondervonden dat het uitstellen van het drukken op de knop deze verstoring verhielp.

Om de N400 nog verder te isoleren van de andere EEG signalen introduceerden we een nieuwe multivariate methode genaamd de spatiële-temporale LCMV beamformer. De performance van de beamformer werd geëvalueerd aan de hang van zowel kunstmatige als echte EEG signalen, alsmede de performance van de traditionele gemiddelde voltage meting en supervised learning technieken, zoals de linear support vector machine (lSVM). We ondervonden dat de performance van de beamformer en lSVM gelijkwaardig waren op grond van correlatie met de daadwerkelijke N400 amplitude, maar dat de beamformer robuuster was tegen invloeden van nabije ERP componenten. De variantie van de output van de beamformer die niet kon worden verklaard door de N400 amplitude was vooral door ongestructureerde ruis, wat een wenselijke eigenschap is tijdens statistische analyse. Het zorgt ervoor dat de onderzoeker zekerheid kan hebben dat de output van de filter daadwerkelijk een maat is van N400 amplitude en niet een andere component.

De combinatie van het experimentele design en de stLCMV beamformer resulteerde in een goede meting van de N400 amplitude. Deze meting werd gebruikt om een studie te doen waarin een lijst van woorden gegroepeerd werd. De groepen waren gebaseerd op de aanname dat woorden die tot dezelfde groep behoren een kleinere N400 amplitude opwekken wanneer ze gepresenteerd worden als woord-paar, dan woorden die tot verschillende groepen behoren. Startend vanuit een ongesorteerde lijst van 14 woorden, lieten we zien dat de N400 component gebruikt kon worden om de woorden te groeperen in hun oorspronkelijke categorieën: 7 Afrikaanse dieren versus 7 meubelen.

Het werk dat gepresenteerd wordt in deze thesis maakt de weg vrij om de N400 te gebruiken als basis voor een meting van de sterkte van semantische relaties tussen woorden, zoals ze gerepresenteerd zijn in het brein.

Curriculum vitae

Personalia

Name Wouter Marijn van Vliet

Birthdate 22-02-1984

Birthplace Apeldoorn, the Netherlands

Nationality Dutch

Address Sammalkalliontie 6A-9, 02210, Espoo, Finland

E-mail w.m.vanvliet@gmail.com

Phone +385 449155561

Current position

Postdoctoral researcher at the Department of Neuroscience and Biomedical Engineering of Aalto University in Finland.

Education

2011-2015 PhD student in biomedical sciences (laboratorium voor neuro- en phychofysiologie, KU Leuven, Belgium

2007-2010 MSc in human machine interaction at the University of Twente, the Netherlands

2003-2007 BSc in computer science at the University of Twente, the Netherlands

Journal articles

[1] Marijn van Vliet, Nikolay Chumerin, Simon De Deyne, Jan Roelf Wiersema, Wim Fias, Gerrit Storms, and Marc M. Van Hulle. “Single-trial ERP component analysis using a spatio-temporal LCMV beamformer.” In: *IEEE Transactions on Biomedical Engineering* (2015), in press.

[2] Marijn van Vliet, Nikolay V. Manyakov, Gert Storms, Wim Fias, Jan R. Wiersema, and Marc M. Van Hulle. “Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs.” In: *PLoS ONE* 9.2 (2014). Ed. by Antoni Rodriguez-Fornells, e87650.

[3] Nikolay V. Manyakov, Nikolay Chumerin, Arne Robben, Adrien Combaz, Marijn van Vliet, and Marc M. Van Hulle. "Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain-computer interfacing." In: *Journal of neural engineering* 10.3 (2013), p. 036011.

[4] Nikolay Chumerin, Nikolay V. Manyakov, Marijn van Vliet, Arne Robben, Adrien Combaz, and Marc M. Van Hulle. "Steady-state visual evoked potential-based computer gaming on a consumer-grade EEG device." In: *IEEE Transactions on Computational Intelligence and AI in Games* 5.2 (2013), pp. 100–110.

Contributions to conference proceedings

[1] Marijn van Vliet, Arne Robben, Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, and Marc M. Van Hulle. "Designing a brain-computer interface controlled video-game using consumer grade EEG hardware." In: *2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*. IEEE, 2012, pp. 1–6.

[2] Elvira Khachatryan, Marijn van Vliet, Simon De Deyne, Gerrit Storms, Hovhannes Manvelyan, and Marc M. Van Hulle. "Amplitude of N400 component unaffected by lexical priming for moderately constraining sentences." In: *2014 4th International Workshop on Cognitive Information Processing (CIP)*. IEEE, 2014, pp. 1–6.

[3] Malypoeur Plong, Kai Shen, Marijn van Vliet, Arne Robben, and Marc Van Hulle. "accurate visual stimulus presentation software for EEG experiments." In: *Proceedings of the First Asian Conference on Information Systems*. Siem Reap, Cambodja, 2012, pp. 1–4.

[4] Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Arne Robben, Marijn van Vliet, and Marc M Van Hulle. "Steady state visual evoked potential based computer gaming–The Maze." In: *Intelligent Technologies for Interactive Entertainment*. Vol. 99. Springer, 2012, pp. 28–37.

[5] Nikolay V. Manyakov, Nikolay Chumerin, Adrien Combaz, Arne Robben, Marijn van Vliet, and Marc M. Van Hulle. "Decoding SSVEP Responses based on Parafac Decomposition." In: *International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)* (2012), pp. 443–447.

[6] Nikolay V. Manyakov, Nikolay Chumerin, Adrien Combaz, Arne Robben, Marijn van Vliet, Patrick A. De Mazière, and Marc M. Van Hulle. "Brain-computer interface research at Katholieke Universiteit Leuven." In: *International Symposium on Applied Sciences in Biomedical and Communication Technologies*. Barcelona, Spain, 2011.

[7] Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Arne Robben, Marijn van Vliet, and Marc M. Van Hulle. “Subject-adaptive steady-state visual evoked potential detection for brain-computer interface.” In: *International Conference on Intelligent Data Acquisition and Advanced Computing Systems*. Prague, Czech Republic, 2011.

[8] Arne Robben, Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Marijn van Vliet, and Marc M. Van Hulle. “Combining object detection and brain computer interfacing: Towards a new way of subject-environment interaction.” In: *2011 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, 2011, pp. 1–6.

[9] NV Manyakov and Nikolay Chumerin. “Decoding phase-based information from steady-state visual evoked potentials with use of complex-valued neural network.” In: *The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2011)*. September. 2011, pp. 369–373.

[10] Nikolay V. Manyakov, Nikolay Chumerin, Adrien Combaz, Arne Robben, Marijn van Vliet, and Marc M. Van Hulle. “Decoding phase-based information from SSVEP recordings: a comparative study.” In: *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. Beijing, China, 2011, pp. 1–6.

[11] Danny Plass-oude Plass-Oude Bos, Matthieu Duvinage, Oytun Oktay, Jaime Delgado Saa, Huseyin G., Ayhan Istanbullu, Marijn van Vliet, Bram van de Laar, and Mannes Poel. “Looking around with your brain in a virtual world.” In: *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain*. IEEE Symposium Series in Computational Intelligence. IEEE, 2011.

[12] Marijn van Vliet, Christian Mühl, Boris Reuderink, and Mannes Poel. “Guessing what's on your mind: using the N400 in brain computer interfaces.” In: *Brain Informatics*. Ed. by Yiyu Yao, Ron Sun, Tomaso Poggio, Jiming Liu, Ning Zhong, and Jimmy Huang. Vol. 6334. Lecture Notes in Computer Science. University of Twente. Springer, 2010, pp. 180–191.

[13] Marijn van Vliet, Alena Neviarouskaya, and Helmut Prendinger. “Opinion elicitation in Second Life.” In: *Intetain*. Ed. by Anton Nijholt, Dennis Reidsma, and Hendry Hondorp. Vol. 9. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Amsterdam: Springer, 2009, pp. 252–257.

Book chapters

[1] Nikolay Chumerin, N. V. Manyakov, Marijn van Vliet, Arne Robben, Adrien Combaz, and Marc M. Van Hulle. “Processing and decoding steady-state visual evoked potentials for brain-computer interfaces.” In: *Digital Image and Signal Processing for Measurement Systems*. 2012, pp. 1–33.

Bibliography

[1] Timothy P. McNamara. “What is semantic priming and why should anyone care about it?” In: *Semantic priming: perspectives from memory and word recognition*. New York, 2005. Chap. 1, pp. 3–7.

[2] J. H. Neely. “Semantic priming effects in visual word recognition: a selective review of current findings and theories.” In: *Basic processes in reading: visual word recognition*. Ed. by D. Besner and G. W. Humphreys. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991. Chap. 9, pp. 264–336.

[3] J. H. Neely. “Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes.” In: *Memory and Cognition* 4.5 (1976), pp. 648–654.

[4] E. Tulving and D. L. Schacter. “Priming and human memory systems.” In: *Science* 247.4940 (1990), pp. 301–306.

[5] Larry R. Squire. “Declarative and nondeclarative memory: multiple brain systems supporting learning and memory.” In: *Journal of Cognitive Neuroscience* 4.3 (1992), pp. 232–243.

[6] Marta Kutas and Kara D. Federmeier. “Electrophysiology reveals semantic memory use in language comprehension.” In: *Trends in Cognitive Sciences* 4.12 (2000), pp. 463–470.

[7] Marta Kutas and Kara D. Federmeier. “Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP).” In: *Annual Review of Psychology* 62 (2011), pp. 621–647.

[8] D. Meyer and R. W. Schvaneveldt. “Facilitation in recognizing pair of words: evidence of a dependence between retrieval operations.” In: *Journal of Experimental Psychology* 90 (1971), pp. 227–234.

[9] R. E. Schuberth, K. T. Spoehr, and D. M. Lane. “Effects of stimulus and contextual information on the lexical decision process.” In: *Memory & Cognition* 9.1 (1981), pp. 68–77.

[10] Phillip J. Holcomb and Helen J. Neville. “Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials.” In: *Language and Cognitive Processes* 5.4 (1990), pp. 281–312.

[11] Guido Orgs, Kathrin Lange, Jan-Henryk Dombrowski, and Martin Heil. “Conceptual priming for environmental sounds and words: an ERP study.” In: *Brain and Cognition* 62.3 (2006), pp. 267–272.

[12] Maria-Teresa Bajo. “Semantic facilitation with pictures and words.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14.4 (1988), pp. 579–589.

[13] Vincent Mm Reid and Tricia Striano. “N400 involvement in the processing of action sequences.” In: *Neuroscience Letters* 433.2 (2008), pp. 93–7.

[14] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor. "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria." In: *Psychophysiology* 37.2 (2000), pp. 127–152.

[15] Shlomo Bentin, Gregory McCarthy, and Charles C. Wood. "Event-related potentials, lexical decision and semantic priming." In: *Electroencephalography and Clinical Neurophysiology* 60 (4 1985), pp. 343–355.

[16] D. E. Keefe and J. H. Neely. "Semantic priming in the pronunciation task: the role of prospective prime-generated expectancies." In: *Memory & Cognition* 18.3 (1990), pp. 289–298.

[17] William S. Maki. "Judgments of associative memory." In: *Cognitive psychology* 54.4 (2007), pp. 319–353.

[18] Tom Heyman, Simon De Deyne, and Gerrit Storms. "Using the letter decision task to examine semantic priming." In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Berlin, 2013, pp. 2542–2547.

[19] M. Lucas. "Semantic priming without association: a meta-analytic review." In: *Psychonomic Bulletin & Review* 7.4 (2000), pp. 618–630.

[20] Annette M. B. de Groot. "Primed lexical decision: combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target." In: *The Quarterly Journal of Experimental Psychology* 36.2 (1984), pp. 253–280.

[21] Allan M. Collins and Elizabeth F. Loftus. "A spreading-activation theory of semantic processing." In: *Psychological Review* 82.6 (1975), pp. 407–428.

[22] John R. Anderson. "Spreading activation theory of memory." In: *Journal of Verbal Learning and Verbal Behaviour* 22 (1983), pp. 261–295.

[23] Timothy P. McNamara and Jeanette Altarriba. "Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions." In: *Journal of Memory and Language* 27.5 (1988), pp. 545–559.

[24] Dorothee J. Chwilla, Herman H. J. Kolk, and Gijsbertus Mulder. "Mediated priming in the lexical decision task: evidence from event-related potentials and reaction time." In: *Journal of Memory and Language* 42.3 (2000), pp. 314–341.

[25] Keith A. Hutchison. "Is semantic priming due to association strength or feature overlap? A microanalytic review." In: *Psychonomic Bulletin & Review* 10.4 (2003), pp. 785–813.

[26] Donna A. Kreher, Phillip J. Holcomb, and Gina R. Kuperberg. "An electrophysiological investigation of indirect semantic priming." In: *Psychophysiology* 43.6 (2006), pp. 550–563.

[27] Colin Brown and Peter Hagoort. “The processing nature of the N400: evidence from masked priming.” In: *Journal of Cognitive Neuroscience* 5.1 (1993), pp. 34–44.

[28] Deacon Deacon, Sean Hewill, Chien-Ming Yang, and Masanouri Nagata. “Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process.” In: *Cognitive Brain Research* 9.2 (2000), pp. 137–146.

[29] Seana Coulson and David Brang. “Sentence context affects the brain response to masked words.” In: *Brain and Language* 113 (2010), pp. 149–155.

[30] P. J. Holcomb and Jonathan Grainger. “On the time course of visual word recognition: an event-related potential investigation using masked repetition priming.” In: *Journal of Cognitive Neuroscience* 18.10 (2006), pp. 1631–1643.

[31] Marianna D. Eddy and Phillip J. Holcomb. “The temporal dynamics of masked repetition picture priming effects: manipulations of stimulus-onset asynchrony (SOA) and prime duration.” In: *Brain Research* 1340 (2010), pp. 24–39.

[32] Mark S. Seidenberg. “Connectionist models of reading.” In: *Oxford Handbook of Psycholinguistics*. Ed. by G. Gaskell. Oxford Univeristy Press, 2007, pp. 235–250.

[33] Tom Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Chang Kai-Min, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. “Predicting human brain activity associated with the meanings of nouns.” In: *Science* 320.January (2008), pp. 1191–1195.

[34] Alexander G. Huth, Shinji Nishimoto, An T. Vu, and Jack L. Gallant. “A continuous semantic space described the representation of thousands of object and action categories across the human brain.” In: *Neuron* 76.6 (2012), pp. 1210–1224.

[35] James L. McClelland and Timothy T. Rogers. “The parallel distributed processing approach to semantic cognition.” In: *Nature Reviews. Neuroscience* 4.April (2003), pp. 310–322.

[36] Sharon L. Thompson-Schill, Kenneth J. Kurtz, and John D. E. Gabrieli. “Effects of semantic and associative relatedness on automatic priming.” In: *Journal of Memory and Language* 38.4 (1998), pp. 440–458.

[37] Ken McRae and Stephen Boisvert. “Automatic semantic similarity priming.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.3 (1998), pp. 558–572.

[38] J. H. Neely, D. E. Keefe, and K. L. Ross. “Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15.6 (1989), pp. 1003–1019.

[39] Donna Coch, Tory Hart, and Priya Mitra. “Three kinds of rhymes: An ERP study.” In: *Brain and Language* 104 (2008), pp. 230–243.

[40] J. Deese. *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press, 1965.

[41] Kevin Lund and Curt Burgess. “Producing high-dimensional semantic spaces from lexical co-occurrence.” In: *Behavior Research Methods, Instruments & Computers* 28.2 (1996), pp. 203–208.

[42] Thomas K. Landauer and Susan T. Dumais. “A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge.” In: *Psychological Review* 104.2 (1997), pp. 211–240.

[43] Magnus Sahlgren, Anders Holst, and Pentti Kanerva. “Permutations as a means to encode order in word space.” In: *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. 2008, pp. 1300–1305.

[44] G. A. Miller. “WordNet: a lexical database for english.” In: *Communication of the ACM* 38.11 (1995), pp. 39–441.

[45] Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. “The University of South Florida free association, rhyme, and word fragment norms.” In: *Behavior Research Methods, Instruments, & Computers* 36.3 (2004), pp. 402–407.

[46] Simon De Deyne and Gert Storms. “Word associations: network and semantic properties.” In: *Behavior Research Methods* 40.1 (2008), pp. 213–231.

[47] Simon De Deyne, Daniel J. Navarro, and Gert Storms. “Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations.” In: *Behavior Research Methods* 45.2 (2013), pp. 480–498.

[48] Douglas L. Nelson, Cathy L. McEvoy, and Simon Dennis. “What is free association and what does it measure?” In: *Memory & Cognition* 28.6 (Sept. 2000), pp. 887–99.

[49] Connie C. Duncan, Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten. “Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400.” In: *Clinical Neurophysiology* 120.11 (2009), pp. 1883–1908.

[50] Marta Kutas and Anders M. Dale. “Electrical and magnetic readings of mental functions.” In: *Cognitive Neuroscience*. Ed. by M. D. Rugg. Hove East Sussex, UK: Psychology Press, 1997. Chap. 7, pp. 197–242.

[51] Micah M. Murray, Denis Brunet, and Christoph M. Michel. “Topographic ERP analyses: a step-by-step tutorial review.” In: *Brain Topography* 20.4 (2008), pp. 249–264.

[52] Paul L. Nunez, Ramesh Srinivasan, Andrew F. Westdorp, Ranjith S. Wijesinghe, Don M. Tucker, Richard B. Silberstein, and Peter J. Cadusch. "EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales." In: *Electroencephalography and Clinical Neurophysiology* 103 (1997), pp. 499–515.

[53] Lucas C. Parra, Chris Alvino, Akaysha Tang, Barak Pearlmutter, Nick Yeung, Allen Osman, and Paul Sajda. "Single-trial detection in EEG and MEG: Keeping it linear." In: *Neurocomputing* 52-54 (2003), pp. 177–183.

[54] Steven J. Luck. *An introduction to the event-related potential technique*. Cambridge, Massachussetts: The MIT Press, 2005. ISBN: 978-0-262-12277-1.

[55] J. Hoormann, M. Falkenstein, P. Schwarzenau, and J. Hohnsbein. "Methods for the quantification and statistical testing of ERP differences across conditions." In: *Behavior Research Methods, Instruments, & Computers* 30.1 (1998), pp. 103–109.

[56] Steven A. Hillyard, Robert F. Hink, Vincent L. Schwent, and Terence W. Picton. "Electrical signs of selective attention in the human brain." In: *Psychophysiology* 182.4108 (1973), pp. 177–180.

[57] Stéphane Dufau, Jonathan GrainGer, and Phillip J. Holcomb. "An ERP investigation of location invariance in masked repetition priming." In: *Cognitive Affective & Behavioural Neuroscience* 8.2 (2008), pp. 222–228.

[58] Jonathan Grainger and Phillip J. Holcomb. "Watching the word go by: on the time-course of component processes in visual word recognition." In: *Language and Linguistics Compass* 3.1 (2009), pp. 128–156.

[59] Krysta Chauncey and Jonathan Grainger. "The effect of stimulus font and size on masked repetion priming." In: *Language and Cognitive Processes* 23.1 (2009), pp. 183–200.

[60] Manuel Martí-Loeches, José. A. Hinojosa, Pilar Casado, Francisco Muñoz, Luis Carretié, Carlos Fernández-Frías, and Miguel A. Pozo. "The recognition potential and repetition effects." In: *International Journal of Psychophysiology* 43.43 (2002), pp. 155–166.

[61] Steven A. Hackley, Marty Woldorff, and Steven A. Hillyard. "Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials." In: *Psychophysiology* 27.2 (1990), pp. 195–208.

[62] B. R. Dunn, D. A. Dunn, M. Languis, and D. Andrews. "The relation of ERP components to complex memory processing." In: *Brain and Cognition* 36.3 (1998), pp. 355–376.

[63] Celeste D. Lefebvre, Yannick Marchand, Gail a. Eskes, and John F. Connolly. "Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task." In: *Clinical Neurophysiology* 116.7 (2005), pp. 1665–1680.

[64] L. A. Farwell and E. Donchin. “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials.” In: *Electroencephalography and Clinical Neurophysiology* 70.6 (1988), pp. 510–523.

[65] Adrien Combaz, Nikolay V. Manyakov, Nikolay Chumerin, Johan A. K. Suykens, and Marc M. Van Hulle. “Feature extraction and classification of EEG signals for rapid P300 mind spelling.” In: *International Conference on Machine Learning and Applications*. IEEE, 2009, pp. 386–391. ISBN: 978-0-7695-3926-3.

[66] N. Chumerin, N. V. Manyakov, A. Combaz, J A. K. Suykens, R. F. Yazicioglu, T. Torfs, P. Merken, H. P. Neves, C. Van Hoof, and Marc M. Van Hulle. “P300 detection based on feature extraction in on-line brain-computer interface.” In: *LNCS KI 2009: Advances in Artificial Intelligence*. Ed. by Mertsching. Vol. 5803/2009. Paderborn, Germany: Springer, 2009, pp. 339–346.

[67] C. C. Duncan-Johnson and E. Donchin. “On quantifying surprise: the variation of event-related potentials with subjective probability.” In: *Psychophysiology* 14.5 (1977), pp. 456–467.

[68] J. Polich. “Updating P300: an integrative theory of P3a and P3b.” In: *Clinical Neurophysiology* 118 (2007), pp. 2128–2148.

[69] Yang Whan Jeon and John Polich. “P3a from a passive visual stimulus task.” In: *Clinical Neurophysiology* 112.12 (2001), pp. 2202–2208.

[70] Marta Kutas and Steven A. Hillyard. “Brain potentials during reading reflect word expectancy and semantic association.” In: *Nature* 307.5947 (1984), pp. 161–163.

[71] L. Osterhout and P. J. Holcomb. “Event-related brain potentials elicited by syntactic anomaly.” In: *Journal of Memory and Language* 31 (1992), pp. 785–806.

[72] Mark Allen, William Badecker, and Lee Osterhout. “Morphological analysis in sentence processing: an ERP study.” In: *Language and Cognitive Processes* 18.4 (2003), pp. 405–430.

[73] Marieke van Herten, Herman H. J. Kolk, and Dorothee J. Chwilla. “An ERP study of P600 effects elicited by semantic anomalies.” In: *Cognitive Brain Research* 22.2 (2005), pp. 241–255.

[74] Marta Kutas, Cyma Van Petten, and Robert Klünder. “Psycholinguistics Electrified II: 1994-2005.” In: *Handbook of Psycholinguistics*. Ed. by M. Traxler and M. A. Gernsbacher. 2nd. New York: Elsevier, 2006. Chap. 4, pp. 659–724.

[75] Marta Kutas and Steven A. Hillyard. “Reading senseless sentences: brain potentials reflect semantic incongruity.” In: *Science* 207.4427 (1980), pp. 203–205.

[76] Marta Kutas and Steven A. Hillyard. “An electrophysiological probe of incidental Semantic association.” In: *Journal of Cognitive Neuroscience* 1.1 (Jan. 1989), pp. 38–49.

[77] Cyma Van Petten and Cyma Van Petten. “Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence.” In: *International Journal of Psychophysiology* 94.3 (2014), pp. 407–419.

[78] O. Hauk and F. Pulvermüller. “Effects of word length and frequency on the human event-related potential.” In: *Clinical Neurophysiology* 115.5 (2004), pp. 1090–1103.

[79] Phillip J. Holcomb, Jonathan Grainger, and Tim O. Rourke. “An electrophysiological study of the effects of orthographic neighborhood size on printed word perception.” In: 1996 (2002), pp. 938–950.

[80] M. Brysbaert. “Age-of-acquisition effects in semantic processing tasks.” In: *Acta Psychologica* 104.2 (2000), pp. 215–226.

[81] Emmanuel Keuleers, Kevin Diependaele, and Marc Brysbaert. “Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords.” In: *Frontiers in psychology* 1.November (2010), p. 174.

[82] Horacio A. Barber, Leun J. Otten, Stavroula-Thaleia Kousta, and Gabriella Vigliocco. “Concreteness in word processing: ERP and behavioral effects in a lexical decision task.” In: *Brain and Language* 125.1 (2013), pp. 47–53.

[83] M. Kutas. “In the company of other words: electrophysiological evidence for single-word and sentence context effects.” In: *Language and Cognitive Processes* 8 (1993), pp. 533–572.

[84] Dorothee J. Chwilla and Herman H. J. Kolk. “Accessing world knowledge: evidence from N400 and reaction time priming.” In: *Cognitive Brain Research* 25.3 (2005), pp. 589–606.

[85] Keith A. Hutchison, David A. Balota, Michael J. Cortese, and Jason M. Watson. “Predicting semantic priming at the item level.” In: *Quarterly journal of experimental psychology (2006)* 61.7 (2008), pp. 1036–66.

[86] Mante S. Nieuwland and Gina R. Kuperberg. “When the truth isn’t too hard to handle: an event-related potential study on the pragmatics of negation.” In: *Psychological Science* 19.12 (2008), pp. 1213–1218.

[87] Lee Osterhout, Michael Bersick, and Richard McKinnon. “Brain potentials elicited by words: word length and frequency predict the latency of an early negativity.” In: *Biological Psychology* 46.2 (1997), pp. 143–168.

[88] G. Ganis, Marta Kutas, and Martin I. Sereno. “The search for ”common sense”: an electrophysiological study of the comprehension of words and pictures in reading.” In: *Journal of Cognitive Neuroscience* 8.2 (1996), pp. 89–106.

[89] Guido Orgs, Kathrin Lange, Jan-Henryk Dombrowski, and Martin Heil. “N400-effects to task-irrelevant environmental sounds: further evidence for obligatory conceptual processing.” In: *Neuroscience Letters* 436.2 (2008), pp. 133–7.

[90] Ela I. Olivares, Jaime Iglesias, and María Antonieta Bobes. “Searching for face-specific long latency ERPs: a topographic study of effects associated with mismatching features.” In: *Cognitive Brain Research* 7.3 (1999), pp. 343–356.

[91] Cyma Van Petten and Barbara J. Luka. “Neural localization of semantic context effects in electromagnetic and hemodynamic studies.” In: *Brain and Language* 97.3 (2006), pp. 279–93.

[92] Ellen F. Lau, Colin Phillips, and David Poeppel. “A cortical network for semantics: (de)constructing the N400.” In: *Nature Reviews. Neuroscience* 9.12 (2008), pp. 920–933.

[93] Burkhard Maess, Christoph S. Herrmann, Anja Hahne, Akinori Nakamura, and Angela D. Friederici. “Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing.” In: *Brain Research* 1096.1 (2006), pp. 163–72.

[94] Johanna Vartiainen, Tiina Parviainen, and Riitta Salmelin. “Spatiotemporal convergence of semantic processing in reading and speech perception.” In: *The Journal of Neuroscience* 29.29 (2009), pp. 9271–80.

[95] Johanna Vartiainen, Mia Liljeström, Miika Koskinen, Hanna Renvall, and Riitta Salmelin. “Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading.” In: *The Journal of Neuroscience* 31.3 (2011), pp. 1048–58.

[96] Burkhard Maess, Christoph S. Herrmann, Anja Hahne, Akinori Nakamura, and Angela D. Friederici. “Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing.” In: *Brain Research* 1096.1 (2006), pp. 163–172.

[97] Ellen Lau, Diogo Almeida, Paul C. Hines, and David Poeppel. “A lexical basis for N400 context effects: evidence from MEG.” In: *Brain and Language* 111.3 (2009), pp. 161–172.

[98] Diana Deacon, Anna Dynowska, Walter Ritter, and Jillian Grose-Fifer. “Repetition and semantic priming of nonwords: implications for theories of N400 and word recognition.” In: *Psychophysiology* 41.1 (2004), pp. 60–74.

[99] J. Bruno Debruille and Louis Renoult. “Effects of semantic matching and of semantic category on reaction time and N400 that resist numerous repetitions.” In: *Neuropsychologia* 47.2 (2009), pp. 506–517.

[100] Holger Hill, Marion Strube, Daniela Roesch-Ely, and Matthias Weisbrod. “Automatic vs. controlled processes in semantic priming–differentiation by event-related potentials.” In: *International Journal of Psychophysiology* 44.3 (2002), pp. 197–218.

[101] Michael Kiang, Iulia Patriciu, Carolyn Roy, Bruce K. Christensen, and Robert B. Zipursky. “Test–retest reliability and stability of N400 effects in a word-pair semantic priming paradigm.” In: *Clinical Neurophysiology* 124.4 (2012), pp. 667–674.

[102] J. Polich. “Semantic categorization and event-related potentials.” In: *Brain and Language* 26.2 (1985), pp. 304–321.

[103] M. Kutas and V. Iragui. “The N400 in a semantic categorization task across 6 decades.” In: *Electroencephalography and clinical neurophysiology* 108.5 (1998), pp. 456–471.

[104] Louis Renoult and J. Bruno Debruille. “N400-like potentials and reaction times index semantic relations between highly repeated individual words.” In: *Journal of Cognitive Neuroscience* 23.4 (2011), pp. 905–922.

[105] Louis Renoult, Xiaoxiao Wang, Jennifer Mortimer, and J. Bruno Debruille. “Explicit semantic tasks are necessary to study semantic priming effects with high rates of repetition.” In: *Clinical Neurophysiology* 123 (2011), pp. 741–754.

[106] Lucas C. Parra, Clay D. Spence, Adam D. Gerson, and Paul Sajda. “Recipes for the linear analysis of EEG.” In: *NeuroImage* 28.2 (2005), pp. 326–41.

[107] Aaron Schurger, Sebastien Marti, and Stanislas Dehaene. “Reducing multi-sensor data to a single time course that reveals experimental effects.” In: *BMC Neuroscience* 14 (2013), p. 122.

[108] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. “Localization of brain electrical activity via linearly constrained minimum variance spatial filtering.” In: *IEEE Transactions on Bio-medical Engineering* 44.9 (1997), pp. 867–880.

[109] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier Science, 2010.

[110] Athina Tzovara, Micah M. Murray, Gijs Plomp, Michael H. Herzog, Christoph M. Michel, and Marzia De Lucia. “Decoding stimulus-related information from single-trial EEG responses based on voltage topographies.” In: *Pattern Recognition* 45.6 (2012), pp. 2109–2122.

[111] E. Donchin and E. Heffley. *Multivariate analysis of event-related potential data: A tutorial review*. Ed. by D. Otto. Government Printing Office, 1978.

[112] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus Robert Müller. “Single-trial analysis and classification of ERP components - a tutorial.” In: *NeuroImage* 56.2 (2011), pp. 814–825.

[113] Steven Lemm, Benjamin Blankertz, Thorsten Dickhaus, and Klaus-Robert Müller. “Introduction to machine learning for brain imaging.” In: *NeuroImage* 56.2 (2011), pp. 387–399.

[114] Jeroen Geuze, Marcel A. J. van Gerven, Jason Farquhar, and Peter Desain. "Detecting Semantic Priming at the Single-Trial Level." In: *PLoS ONE* 8.4 (2013). Ed. by Emmanuel Andreas Stamatakis, e60377.

[115] Johannes Muller-Gerking, Gert Pfurtscheller, and Henrik Flyvbjerg. "Designing optimal spatial filters for single-trial EEG classification in a movement task." In: *Clinical Neurophysiology* 110 (1999), pp. 787–798.

[116] Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert. "xDAWN algorithm to enhance evoked potentials: application to brain-computer interface." In: *IEEE Transactions on Bio-medical Engineering* 56.8 (2009), pp. 2035–2043.

[117] Gabriel Pires, Urbano Nunes, and Miguel Castelo-Branco. "Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis." In: *Journal of Neuroscience Methods* 195.2 (2011), pp. 270–81.

[118] Ke Yu, Kaiquan Shen, Shiyun Shao, Wu Chun Ng, Kenneth Kwok, and Xiaoping Li. "A spatio-temporal filtering approach to denoising of single-trial ERP in rapid image triage." In: *Journal of Neuroscience Methods* 204.2 (2012), pp. 288–95.

[119] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus Robert Müller. "Single-trial analysis and classification of ERP components - A tutorial." In: *NeuroImage* 56.2 (2011), pp. 814–825.

[120] M. Kutas, G. McCarthy, and E. Donchin. "Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time." In: *Science* 197 (1977), pp. 792–795.

[121] Judith M. Ford, Richard C. Mohs, Adolf Pfefferbaum, and Bert S. Kopell. "On the utility of P3 latency and RT for studying cognitive processes." In: *Progress in Brain Research* 54 (1980), pp. 661–667.

[122] Yaqub Jonmohamadi, Govinda Poudel, Carrie Innes, and Daniel Weiss. "Comparison of beamformers for EEG source signal reconstruction." In: *Biomedical Signal Processing and Control* 14 (2014), pp. 175–188.

[123] Anne Cutler. "Making up materials is a confounded nuisance, or: will we be able to run any psycholinguistic experiments at all in 1990?" In: *Cognition* 10 (1981), pp. 65–70.

[124] R. McNamee. "Regression modelling and other methods to control confounding." In: *Occupational and Environmental Medicine* 62.7 (2005), pp. 500–506, 472.

[125] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-113. 2013.

[126] Alexander M. Chan, Eric Halgren, Ksenija Marinkovic, and Sydney S. Cash. "Decoding word and category-specific spatiotemporal representations from MEG and EEG." In: *NeuroImage* 54.4 (2011), pp. 3028–39.

[127] Mika Koivisto and Antii Revonsuo. “Cognitive representations underlying the N400 priming effect.” In: *Cognitive Brain Research* 12.3 (2001), pp. 487–490.

[128] Yusuke Takeda, Kentaro Yamanaka, and Yoshiharu Yamamoto. “Temporal decomposition of EEG during a simple reaction time task into stimulus- and response-locked components.” In: *NeuroImage* 39.2 (2008), pp. 742–754.

[129] Penny F. Mitchell, Sally Andrews, and Philip B. Ward. “An event-related potential study of semantic congruity and repetition in a sentence-reading task: effects of context change.” In: *Psychophysiology* 30.5 (1993), pp. 496–509.

[130] Emmanuel Keuleers, Marc Brysbaert, and Boris New. “SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles.” In: *Behavior Research Methods* 42.3 (2010), pp. 643–650.

[131] R. J. Croft and R. J. Barry. “Removal of ocular artifact from the EEG: a review.” In: *Clinical Neurophysiology* 30.1 (2000), pp. 5–19.

[132] D. H. Brainard. “The Psychophysics Toolbox.” In: *Spatial Vision* 10 (1997), pp. 433–436.

[133] Travis E. Oliphant. “Python for scientific computing.” In: *Computing in Science Engineering* 9.3 (2007), pp. 10–20.

[134] J.D. Hunter. “Matplotlib: a 2D graphics environment.” In: *Computing in Science Engineering* 9.3 (2007), pp. 90–95.

[135] R. H. Baayen, D. J. Davidson, and Bates D. M. “Mixed-effects modeling with crossed random effects for subjects and items.” In: *Journal of Memory and Language* 59 (2008), pp. 390–412.

[136] F. E. Satterthwaite. “An approximate distribution of estimates of variance components.” In: *Biometrics* 2.6 (1946), pp. 110–114.

[137] Walter Ritter, Richard Simson, and Herbert G. Vaughan Jr. “Association cortex potentials and reaction time in auditory discrimination.” In: *Electroencephalography and Clinical Neurophysiology* 33 (1972), pp. 547–555.

[138] S. K. Jankelowitz and J. G. Colebatch. “Movement-related potentials associated with self-paced, cued and imagined arm movements.” In: *Experimental Brain Research* 147.1 (2002), pp. 98–107.

[139] H. Shibasaki, G. Barrett, E. Halliday, and A. M. Halliday. “Components of the movement-related cortical potential and their scalp topography.” In: *Electroencephalography and Clinical Neurophysiology* 49.3-4 (1980), pp. 213–226.

[140] C. D. Woody. “Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals.” In: *Medical and Biological Engineering* 5 (1967), pp. 539–553.

[141] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. "Analysis and visualization of single-trial event-related potentials." In: *Human Brain Mapping* 14.3 (2001), pp. 166–185.

[142] Thomas F. Münte, Thomas P. Urbach, Emrah Düzel, and Marta Kutas. "Event-related brain potentials in the study of human cognition and neuropsychology." In: *Handbook of Neuropsychology*. Ed. by F. Boller, J. Grafman, and G. Rizzolatti. 2nd. Vol. 1. Elsevier Science, 2000. Chap. 7, pp. 1–97. ISBN: 978-0-444-50358-9.

[143] T. C. Handy. *Event-related potentials: a methods handbook*. Cambridge: MIT Press, 2004, p. 416. ISBN: 9780262083331.

[144] Megan A. Boudewyn, Debra L. Long, and Tamara Y. Swaab. "Cognitive control influences the use of meaning relations during spoken sentence comprehension." In: *Neuropsychologia* 50.11 (2012), pp. 2659–68.

[145] Michael Dambacher, Reinhold Kliegl, Markus Hofmann, and Arthur M. Jacobs. "Frequency and predictability effects on event-related potentials during reading." In: *Brain research* 1084.1 (2006), pp. 89–103.

[146] Kara D. Federmeier, Marta Kutas, and Rina Schul. "Age-related and individual differences in the use of prediction during language comprehension." In: *Brain and language* 115.3 (2010), pp. 149–61.

[147] Sarah Laszlo and Kara D. Federmeier. "The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects." In: *Psychophysiology* 48 (2010), pp. 176–186.

[148] Emmeke Aarts, Matthijs Verhage, Jesse V Veenvliet, Conor V. Dolan, and Sophie van der Sluis. "A solution to dependency: using multilevel analysis to accommodate nested data." In: *Nature Neuroscience* 17.4 (2014), pp. 491–6.

[149] Herbert H. Clark. "The language-as-fixed-effect fallacy: a critique of language statistics in psychological research." In: *Journal of Verbal Learning and Verbal Behaviour* 12.4 (1973), pp. 335–359.

[150] Julian J. Faraway. *Extending the linear model with R*. Boca Raton, FL: Chapman & Hall/CRC, 2006. ISBN: 1-56488-424-8.

[151] Brady T. West, Kathleen B. Welch, and Andrzej T. Galecki. *Linear mixed models: a practical guide using statistical software*. Boca Raton, FL: Taylor & Francis Group, LLC, 2007.

[152] Helen Vossen, Gerard van Breukelen, Hermie Hermens, Jim van Os, and Richel Lousberg. "More potential in statistical analyses of event-related potentials: a mixed regression approach." In: *International journal of methods in psychiatric research* 20.3 (2011), e56–68.

[153] Ernst Niedermeyer and F. H. Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[154] E Donchin. “Discriminant analysis in average evoked response studies: the study of single trial data.” In: *Electroencephalography and clinical neurophysiology* 27 (1969), pp. 311–314.

[155] Cyril R. Pernet, Paul Sajda, and Guillaume a. Rousselet. “Single-trial analyses: Why bother?” In: *Frontiers in Psychology* 2.November (2011), pp. 1–2.

[156] B. D. Van Veen and K. M. Buckley. “Beamforming: a versatile approach to spatial filtering.” In: *IEEE ASSP Magazine* 5.April (1988), pp. 4–24.

[157] Moritz Grosse-Wentrup, Christian Liefhold, Klaus Gramann, and Martin Buss. “Beamforming in noninvasive brain-computer interfaces.” In: *IEEE Transactions on Bio-medical Engineering* 56.4 (2009), pp. 1209–19.

[158] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. “Shrinkage algorithms for MMSE covariance estimation.” In: *IEEE Transactions on Signal Processing* 58.10 (2010), pp. 5016–5029.

[159] Denis A. Engemann and Alexandre Gramfort. “Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals.” In: *NeuroImage* 108 (2015), pp. 328–342.

[160] Juliane Schäfer and Korbinian Strimmer. “A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics.” In: *Statistical Applications in Genetics and Molecular Biology* 4.1 (2005), Article32.

[161] Marijn van Vliet, Nikolay V. Manyakov, Gert Storms, Wim Fias, Jan R. Wiersema, and Marc M. Van Hulle. “Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs.” In: *PLoS ONE* 9.2 (2014). Ed. by Antoni Rodriguez-Fornells, e87650.

[162] Christopher J. C. Burges. “A Tutorial on Support Vector Machines for Pattern Recognition.” In: *Data Mining and Knowledge Discovery*. NetGames ’06 2.2 (1998). Ed. by Usama Fayyad, pp. 121–167.

[163] Alexandre Gramfort, Théodore Papadopoulo, Emmanuel Olivi, and Maureen Clerc. “Forward field computation with OpenMEEG.” In: *Computational Intelligence and Neuroscience* 2011 (2011), p. 923703.

[164] Emmanuel Keuleers, Marc Brysbaert, and Boris New. “SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles.” In: *Behavior research methods* 42.3 (2010), pp. 643–650.

[165] Marc Brysbaert, Michaël Stevens, Simon De Deyne, Wouter Voorspoels, and Gert Storms. “Norms of age of acquisition and concreteness for 30,000 Dutch words.” In: *Acta psychologica* 150 (2014), pp. 80–4.

[166] Alexandre Barachant and Marco Congedo. *A plug & play P300 BCI using information geometry*. Tech. rep. Grenoble, France: GIPSA-lab, CNRS, Grenoble Universities, 2014, pp. 1–9. arXiv: 1409.0107v1. URL: http://arxiv.org/abs/1409.0107.

[167] Jan Kybic, Maureen Clerc, Olivier Faugeras, Renaud Keriven, and Théo Papadopoulo. "Generalized head models for MEG/EEG: boundary element method beyond nested volumes." In: *Physics in medicine and biology* 51.5 (2006), pp. 1333–1346.

[168] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: machine learning in python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[169] Alexandre Barachant. *pyRiemann v0.2.2*. June 2015. DOI: 10.5281/zenodo.18982. URL: http://dx.doi.org/10.5281/zenodo.18982.

[170] Maximilian Nickel. *Scikit-tensor*. https://github.com/mnick/scikit-tensor. June 2015.

[171] R Development Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2008. URL: http://www.R-project.org.

[172] Douglas Bates, Martin Maechler, Benjamin M. Bolker, and Steven Walker. *Fitting Linear Mixed-Effects Models using lme4*. ArXiv e-print; in press, *Journal of Statistical Software*. 2015. URL: http://arxiv.org/abs/1406.5823.

[173] Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0-6. 2014. URL: http://CRAN.R-project.org/package=lmerTest.

[174] S. De Deyne and G. Storms. "Word associations: norms for 1,424 Dutch words in a continuous task." In: *Behavior Research Methods* 40.1 (2008), pp. 198–205.

[175] Qingfei Chen, Chun Ye, Xiuling Liang, Bihua Cao, Yi Lei, and Hong Li. "Automatic processing of taxonomic and thematic relations in semantic priming–differentiation by early N400 and late frontal negativity." In: *Neuropsychologia* 64C (2014), pp. 54–62.

[176] Lei Wang, Qingguo Ma, Zhaofeng Song, Yisi Shi, Yi Wang, and Lydia Pfotenhauer. "N400 and the activation of prejudice against rural migrant workers in China." In: *Brain Research* 1375 (2011), pp. 103–110.

[177] Mireille Besson, Marta Kutas, and Cyma Van Petten. "An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences." In: *Journal of Cognitive Neuroscience* 4.2 (1992), pp. 132–149.

[178] Patrick Doreian. "An intuitive introduction to blockmodeling with examples." In: *Bulletin de Méthodologie Sociologique* 61.1 (1999), pp. 5–34.

[179] Jan W. Van Strien, Rogier E. Hagenbeek, Cornelis J. Stam, Serge A. R. B. Rombouts, and Frederik Barkhof. "Changes in brain electrical activity during extended continuous word recognition." In: *NeuroImage* 26 (2005), pp. 952–959.

[180] Niels Birbaumer. “Breaking the silence: brain-computer interfaces (BCI) for communication and motor control.” In: *Psychophysiology* 43.6 (2006), pp. 517–532.

[181] Olivier Ledoit and Michael Wolf. “Honey, I shrunk the Sample covariance matrix.” In: *The Journal of Portfolio Management* 30.4 (2004), pp. 110–119.

[182] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. “Mixed Membership Stochastic Blockmodels.” In: *Journal of Machine Learning Research* 9 (2008), pp. 1981–2014.